PERFORMANCE ON THE EXPANDED TIME BEARING PLOT AS A FUNCTION OF BEARING ACCURACY

by

Gary M. Olson, LT, MSC, USN
and
Kevin Laxar

NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY
REPORT NUMBER 716

Bureau of Medicine and Surgery, Navy Department
Research Work Unit MF51.524.004-2002DX5G.01

Reviewed and Approved by:

Charles F. Gell

Charles F. Gell, M.D., D.Sc.(Med)
Scientific Director
NavSubMedRschLab

Approved and Released by:

R. L. Sphar

R. L. Sphar, CDR MC USN
Officer in Charge, Acting
NavSubMedRschLab

SUMMARY PAGE

THE PROBLEM

To evaluate the effects of various kinds and amounts of statistical noise in raw sonar bearings on performance in the expanded time bearing plot.

FINDINGS

Performance for both human plotters and a simple mathematical curve fitting routine was affected similarly by the characteristics of the noise, but overall performance of the mathematical routine was superior on simple problems while humans were better on more complicated problems and at the ends of problems.

APPLICATION

This research should provide empirical support for the development of interactive systems which exploit the special abilities of both human and automated procedures in a complicated information processing task.

ADMINISTRATIVE INFORMATION

This investigation was conducted as part of Bureau of Medicine and Surgery Research Work Unit MF51.524.004-2002DX5G - Man as an Information Processor.

The present report was approved for publication on 26 June 1972. It is Report No. 1 on the indicated Work Unit and has been designated as Naval Submarine Medical Research Laboratory Report No. 716.

PUBLISHED BY THE NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY

ABSTRACT

Two experiments analyzed the effects of statistical noise in raw sonar bearings on performance in a laboratory version of the expanded time bearing plot. Accuracy of faired bearings and bearing rate estimates were taken as the measures of performance. Greater amounts of noise led to poorer performance, but these decrements were smaller when the noise was random than when it was correlated. Human performance was contrasted with that of an orthogonal polynomial curve-fitting routine designed to do the same task. The mathematical routine was affected by the noise in the same way as humans were. However, on simple plots the mathematical routine provided superior solutions while on curves of more complex shapes or at the ends of curves humans were superior. Thus, in certain situations the human's perceptual and cognitive abilities gave him a distinct advantage over the mathematical routine.

PERFORMANCE ON THE EXPANDED TIME BEARING PLOT AS A FUNCTION OF BEARING ACCURACY

INTRODUCTION

One of the functions of the submarine's fire control party is that of determining a potential target's course, range, and speed on the basis of passive sonar bearing inputs. Elaborate techniques have been developed to accomplish this, using both human and computer information processing capabilities. The expanded time bearing plot, one of several manual sonar plots in the fire control system, is of central importance in this analysis since its outputs are used by several key stations in the system. In this plot the raw compass bearings obtained from sonar are plotted against elapsed time since initial contact. Then a curve is sketched which captures the central tendency of the plotted points, and from this curve the rate of change of the bearings is estimated using a specially designed bearing rate template. The bearing rate is an especially critical input to the fire control party's analysis of target motion.

Changes in current technology and operational goals have made it possible to track targets at great ranges, and as a result the raw bearings transmitted from sonar may be subject to greater distortion than previously. Further, since the bearing rates at greater ranges are smaller, the effects of distortions will be correspondingly larger. These distortions show up in the time bearing plot as dispersions of the plotted points around some centrally tending curve. The plotter attempts to recapture this by fairing a curve through the plotted points. The investigations reported here focussed on the effects of distortions in the bearings on time bearing plotter performance, under conditions simulating long-range, low bearing rate contacts.

The noise (we shall hereafter use this information theoretic term to refer to the distortions in the bearings) can have varying characteristics. This noise can vary in its magnitude, and in fact when it gets too great, the sonar operator will switch his mode of reporting until he is once more confident he can track the target continuously. The noise can also vary in kind. It may be essentially random, as though it had been generated by a stochastically independent random noise generator. Here the magnitude and direction of a particular raw bearing's deviation from its corresponding actual bearing is independent of any other bearing's deviation. A more probable kind of noise in operational situations is that which is stochastically nonindependent or correlated. With such noise the magnitude and direction of any given raw bearing's deviation is related to or correlated with that of other bearings near it in time. This would appear as a slow drift in the bearings, first to one side and then to the other, around the actual bearings. This type of error is potentially more serious, since it is less noticeable on visual inspection (the correlation tends to smooth the curve) and could lead to systematic biases in the estimates of faired bearings and bearing rates. Either kind of noise

could conceal discontinuities in the time bearing curve which are indicative of target maneuvers or could lead the plotter to falsely report a target maneuver when none existed.

Previous investigations of time bearing performance were concerned with either the older vertical plotting system or with noise-free bearings. The two experiments reported here examined performance on the expanded time bearing plot with variations in both the level and kind of noise in the raw bearings. A simple version of the operational time bearing plot was designed for experimental purposes in order to see the best a human plotter can do under varying conditions of signal degradation. All problems simulated long-range contacts with low bearing rates. Thus, plotters were never forced to change their plotting scales during a problem. Further, no maneuvers by own ship or by target were simulated and the plotter worked at his own pace rather than in real time. Thus, performance under these conditions could be interpreted as the best a plotter could do given the signal characteristics, since the additional factors involved in an actual operational plot would only serve to magnify the performance errors found here. The details of the plots used will be discussed in the next section.

METHOD

The two experiments shared a general methodology, and this will be described separately from their unique design characteristics.

Description of the Task

A special experimental version of the expanded time bearing plot was created which departed from actual operational plots in several ways. All plotting was done on a single sheet of 11 x 16-1/2 in. graph paper ruled every tenth of an inch. Several constraints had to be introduced because the smaller plotting sheet was used. Each problem simulated a portion of a time bearing plot in which a contact had already been made and was being tracked. This contact was followed in the experimental problem for 15 minutes of hypothetical time, these times being labeled from 0:00 to 15:00 for plotting purposes but not necessarily corresponding to the first 15 minutes of contact. Similarly, the range of possible bearings was restricted so that the entire plot could be done on a single sheet of plotting paper without any change of scale (a one degree per inch by one minute per inch scale was used throughout). Thus, for any given problem the raw bearings were not allowed to change by more than 10.5 degrees in the 15 minutes of hypothetical time. In addition, there were no changes in the direction of the bearing rate (right to left or vice versa), the target was assumed to be on a constant course with no target or own ship maneuvers, and the bearing rate was not allowed to get very large (generally being less than one degree per minute in these problems).

Several procedural changes were introduced for experimental purposes which also differed from procedures used in operational tasks. Subjects were presented all the raw bearings at once on a computer printed sheet and allowed to

plot at their own pace, so that although a 15-minute time slice was examined, plotting was not done in real time. Similarly, they were asked to produce faired bearings and estimates of the bearing rate at prespecified intervals and write this in appropriate spaces on the same computer printout. Raw bearings were presented every 30 seconds from time 0:00 to 15:00 inclusive, yielding a total of 31 points to be plotted. Faired bearings were requested every whole minute from time 1:00 to time 14:00 for a total of 14 faired bearings, while estimates of the bearing rate were requested for alternate whole minutes from times 2:00 to 12:00 inclusive for a total of six bearing rate estimates.

Subject to these constraints, a set of twelve standard problems was created for use in these experiments. Each problem was derived either from some appropriate mathematical function or from a set of line segments which yielded a curve approximating the characteristics of curve segments from operational plots. The set of twelve problems are shown in Figure 1.

Independent Variables

These studies were designed to investigate the effects of noise in the raw bearings upon subjects' ability to estimate actual bearings and bearing rates on the basis of the time bearing plot. Two aspects of such noise were examined: (1) The <u>level</u> of noise was manipulated by controlling the standard deviation of the noise in a special computer program which generated the pseudo-random degradation. (See Appendix A for details on the procedure used to generate the noise.) Examples of <u>low</u>, <u>medium</u>, and <u>high</u> levels of noise are shown in Figure 2. The <u>kind</u> of noise was also manipulated, and this referred to correlations between the signed magnitude of the degradation for successive raw bearings. In the <u>random</u> noise condition the magnitude of the error introduced for any given bearing was statistically independent of the magnitude of the error for any other bearing. This is in contrast to the <u>correlated</u> noise condition where the magnitude of the degradation for any particular bearing is related to or correlated with the magnitude of the degradations for bearings near it. Figure 2 also illustrates this difference.

A third independent variable was <u>time</u>, or the place in the problem where the subject provided an estimate. There were fourteen bearing estimates (minutes 1-14) and six bearing rate estimates (even minutes from 2 to 12). Note that time refers to the hypothetical time represented by the ordinate of each plot, <u>not</u> actual elapsed time in the experimental situation.

Dependent Variables

Subjects' performance was compared with two kinds of criteria. One comparison involved contrasting their estimates with the actual bearings and bearing rates obtained from the mathematical functions characterizing each of the 12 standard problems (see Figure 1). This provided a measure of how well the subjects could recover the "true" state of affairs that was obscured by the statistical noise. The second comparison involved contrasting subjects' estimates with the best estimates

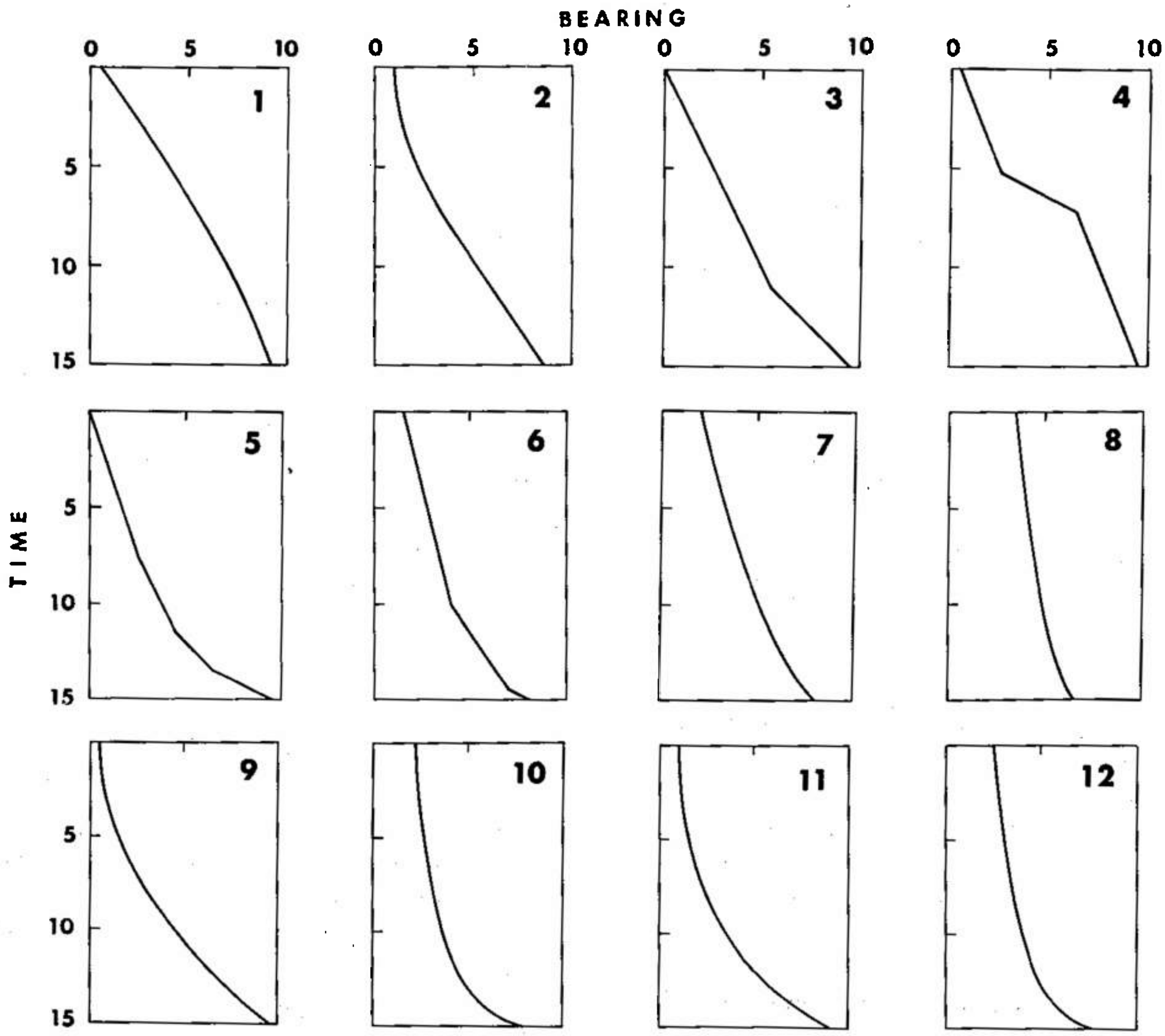

Fig. 1. The set of time bearing problems, without degradation, showing only right bearing rates.

produced by a mathematical curve-fitting technique. This comparison provided an indication of how subjects' performance deviated from that of an analytic curve-fitting procedure.

Two types of measures were obtained for both bearing and bearing rates. The "true" underlying parameter was subtracted from a subject's estimate of a parameter, and this signed score was a measure of algebraic error. It should be pointed out that algebraic errors were computed by subtracting the actual bearing from the estimated bearing. In other words, a negative value indicates that the estimated bearing was smaller than the actual bearing while a positive value means the estimated bearing was larger. Greater than and less than were defined with respect to

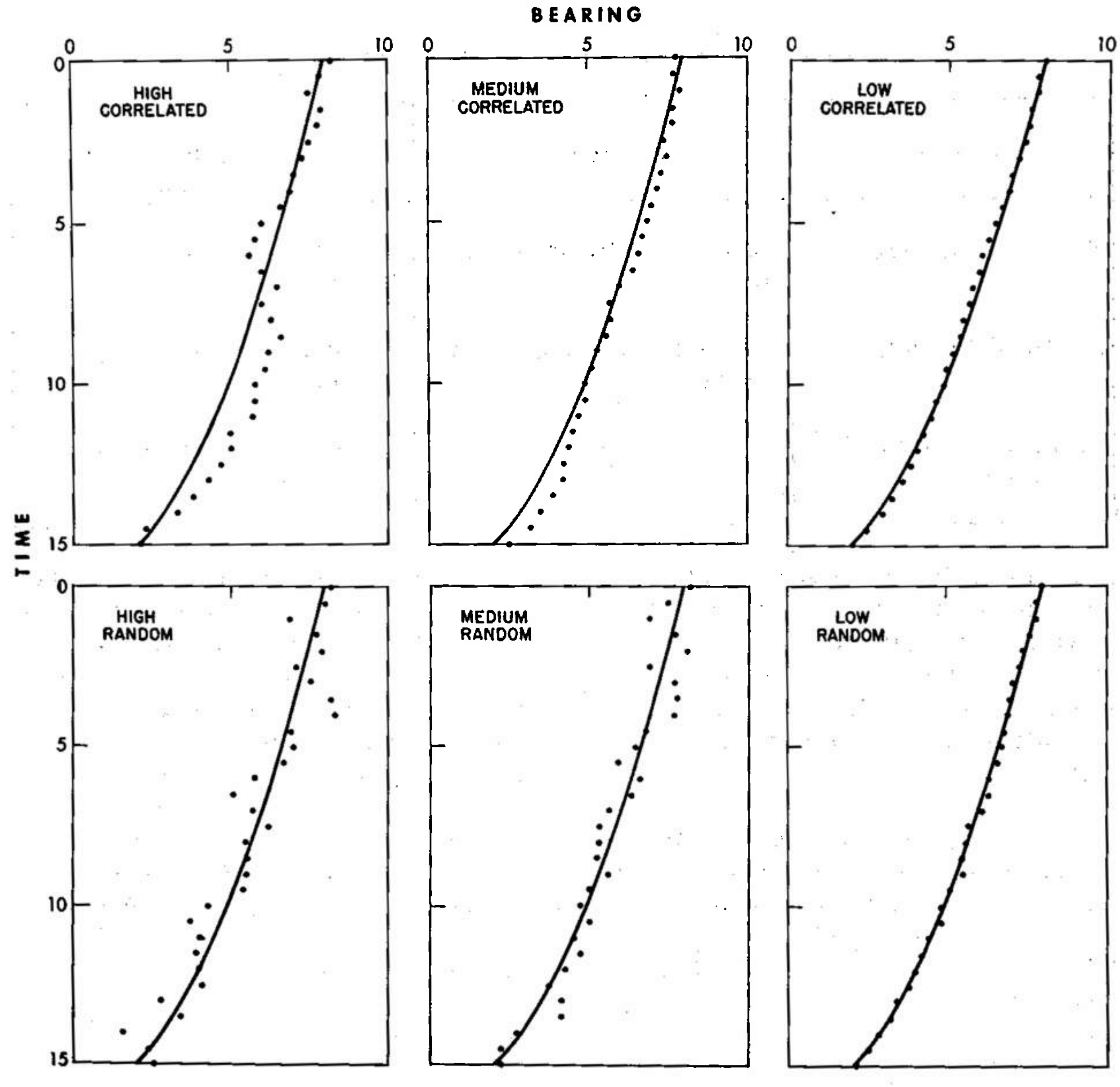

Fig. 2. Examples of the three levels and two kinds of bearing degradation (noise), with the actual underlying curve shown.

the value of the actual bearing, with the convention that if an estimate were just below 360° and the actual just above 000°, or vice versa, the estimate was first translated to the scale of the actual by adding or subtracting 360° as appropriate. The absolute value of this score was a measure of absolute error. A third measure was used for bearing rates only. The absolute error was divided by the true underlying bearing rate to yield a proportion absolute error. These various measures are summarized in Table I.

General Procedure

In both experiments, subjects were scheduled as groups for two-hour

Table 1. Summary of Dependent Variables

Measure	Bearing	Bearing Rate
1. Algebraic error	Subject's estimate - actual bearing	Subject's estimate - actual bearing rate
2. Absolute error	Absolute value (algebraic error)	Absolute value (algebraic error)
3. Proportion Absolute Error	----------------	Absolute error/Actual bearing rate

sessions during which they completed four time bearing plots. All subjects started a given plot at the same time, but were allowed to work at their own speed. A new problem was not started until all subjects had finished the previous one.

At the beginning of a session each subject was given a pencil and a flexible, transparent bearing rate template identical to those used in the fleet. When everyone was ready to begin a new plot the experimenter passed out blank plotting paper and an individualized computer print-out. Subjects had been instructed to do each plot in the following manner: enter the date and subject identification number on both sheets, label the axes of the graph (scales for labeling purposes were given on the print-out), plot the raw bearings, fair a smooth curve through the raw bearings, enter the required faired bearings on the computer print-out, and estimate the required bearing rates and enter these on the print-out.

This sequence of operations obviously differs from the sequence one would follow doing a plot in real time at sea. Nonetheless, by controlling a number of extraneous and complicating influences found in operational versions the experimental procedure allowed a relatively pure assessment of the effects of the independent variables. It can safely be assumed that any effects attributed to the independent variables in these experiments would be magnified under the more complicated situation at sea.

Design

Experiment I. Twelve subjects were each given 12 plots to do, four a session for each of three two-hour sessions. The three experimental sessions had been preceded by two two-hour training and practice sessions in which subjects had been taught how to do time bearing plots and had been given practice in the experimental procedure until the experimenter was satisfied all subjects were competent in all aspects of the task.

Three levels of noise (high, medium, and low) were combined with two kinds of noise (correlated, random) and two directions of bearing rate (right, left). Kind of noise was a between-subject variable; level of noise and direction of bearing rate were within-subject variables.

Each subject received all 12 standard problems (see Figure 1), with level of noise and direction of bearing rate assigned by means of a modified Latin squares procedure. Thus, any given subject had four problems at each of the three levels of noise, half of these presented with right bearing rates, half with left. The Latin squares procedure ensured that the assignment of values of the within-subject variables was counterbalanced across subjects and problem types. Independent random permutations were used to establish the presentation order of the twelve plots for each subject.

Experiment II. This experiment was run as part of a comprehensive 30-day sonar confinement study, and constituted one set of performance measures among a wide variety of performance, perceptual, and physiological measures used in that study. Two complete replications of a within-subject study of time bearing performance were run, one replication early and another late in the noise portion of the sonar habitability study. In each replication a unique series of 12 plots was prepared for each subject. Four plots were done on each of three two-hour evening sessions during a given replication. A three-hour training and practice session was held prior to the first replication.

Two levels of noise (high, medium) were combined factorially with two kinds of noise (correlated, random) and applied to three problem types to generate the twelve plots presented to each subject. Direction of bearing rate was counterbalanced across these. Each subject saw three different problems selected from those shown in Figure 1. (Two subjects saw Problems 6, 7, and 9, three subjects saw Problems 2, 8, and 12, and four subjects saw Problems 1, 10, and 11. Problems 3, 4, and 5 were not used in Experiment II.) Different randomizations were used to create each new set of raw bearings, as in Experiment I. For a particular subject the two sets of 12 plots he saw in the two replications differed only in the randomization used to generate the set of bearings and the random permutation used to establish the order of presentation.

Subjects

Experiment I. Twelve Navy enlisted men waiting to begin Submarine School served as subjects. None of them had prior experience with the expanded time bearing plot or with submarine fire control problems.

Experiment II. All nine subjects who participated in the 30-day confinement study were used in this experiment. Four subjects were civilians from the local community who were paid for their participation; none had prior experience with submarine operations or the Naval service in general. The remaining five subjects were Naval enlisted men. All were experienced sonar technicians familiar with the expanded time bearing

plot and submarine fire control problems, but only one man reported having extensive experience.

RESULTS AND DISCUSSION

For each dependent variable, the appropriate set of scores was obtained for each time segment of each of the problems. Special difficulties were encountered in the analysis of Problem 4 in Experiment I. Therefore, this problem was excluded from the main analysis and it will be discussed in a separate section. For a given subject, an average score on each dependent variable for each of the three levels of noise and each of the time segments (six for bearing rate, 14 for bearing) constituted his input to the analyses about to be reported. Preliminary analyses indicated there were no effects due to direction of the bearing rate, so this variable was ignored in all subsequent analyses.

Similar average scores were obtained for each subject in Experiment II. Once again direction of bearing rate was ignored since preliminary analyses indicated it had no effect on performance. Since there were no differences in performance for the early and the late sessions, these data were also combined. Thus, for each subject, a set of scores was obtained for each dependent variable by averaging over problems and replications for each combination of level (high vs. medium) and kind (random and correlated) of noise at each of the time samplings.

The data for each dependent variable in Experiment I were subjected to a one-between (two levels of kind of noise), two-within (three levels of amount of noise, either six or 14 levels of time of estimate) mixed design analysis of variance (Winer, 1971[1], sec. 7.3). Those from Experiment II were subjected to a three factor (two levels of kind of noise, two levels of amount of noise, either six or 14 levels of time of estimate) completely-crossed within-subject analysis of variance (Winer, 1971[1], sec. 7.5).

Estimation of Bearings

Figures 3 and 4 summarize the performance of subjects at estimating bearings, and are separated according to the independent and dependent variables. The lower panels of these figures show that the algebraic errors tended to remain fairly close to zero in both experiments. The deviations from zero appear to be unsystematic. Thus, there was no overall bias in the directions of the errors subjects made with respect to the actual underlying bearings. Furthermore, the analyses of variance for algebraic errors in estimating bearings revealed that there were no significant effects for any of the independent variables or their interactions in either experiment. In sum, the average signed deviation of the subjects' estimates did not depart systematically from zero and were not influenced by any of the independent variables (kind, level, time).

The upper panels of Figures 3 and 4 show the data for absolute deviations. This measure is an index of how much on the average an estimated bearing deviated from the actual bearing, regardless of the direction of deviation.

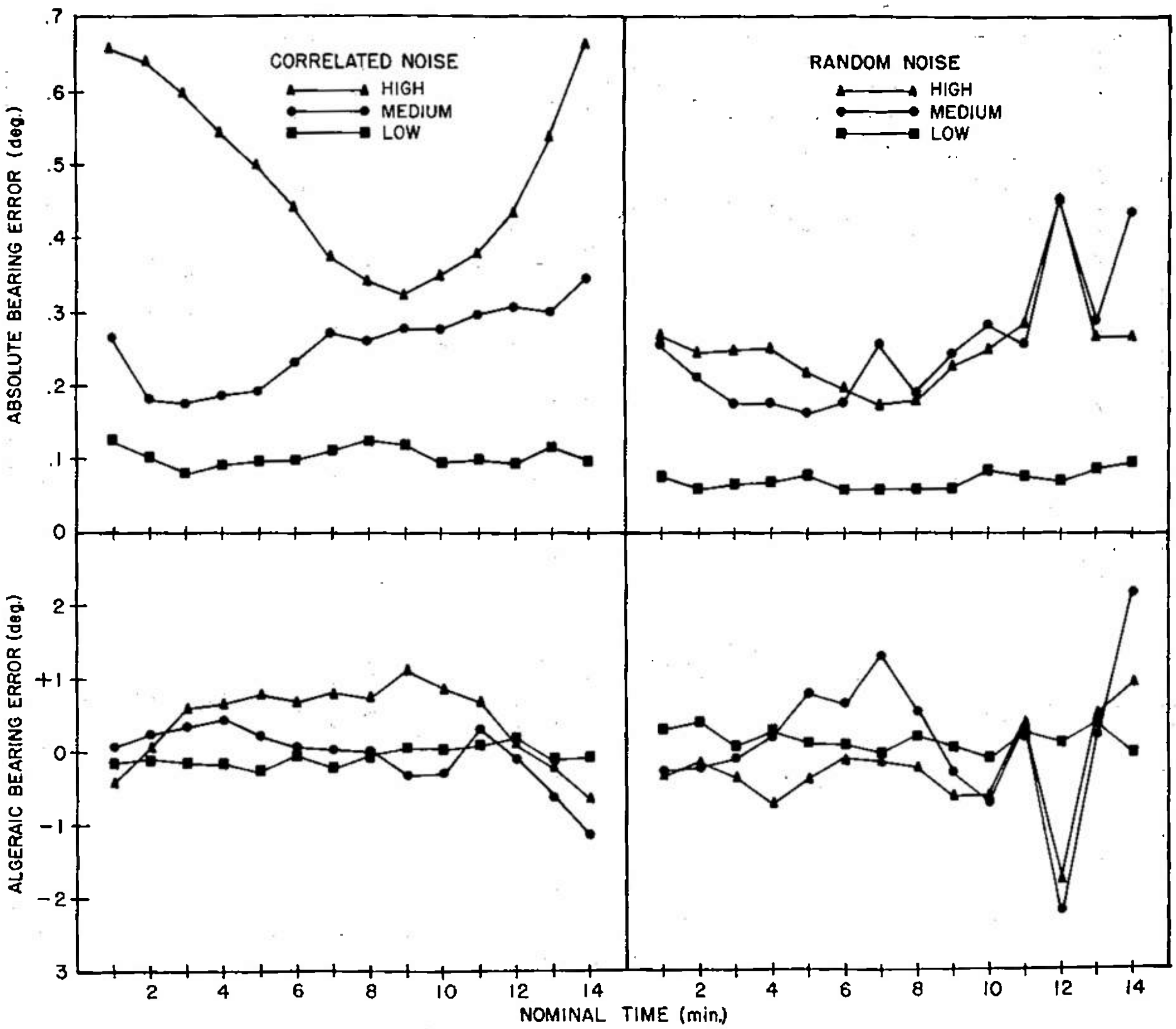

Fig. 3. Error in bearing estimates for Experiment I at three levels of noise.

These panels show that there were systematic differences in the size of these deviations, and statistical analyses confirmed that a number of these differences were reliable. In both experiments the average absolute deviations were greater when the noise was correlated than when it was random (Exp. I: $F_{1,10} = 8.16$, $p < .05$; Exp. II: $F_{1,8} = 105.49$, $p < .001$). It should be recalled that in Experiment I this was a between-subject variable, while in Experiment II it was a within-subject one. In either case performance was significantly affected. Similarly, the average absolute deviation increased as the level of noise was increased (Exp. I: $F_{2,20} = 77.15$, $p < .001$; Exp. II: $F_{1,8} = 23.78$, $p < .01$). In Experiment I these two factors had a significant interaction ($F_{2,20} = 15.23$, $p < .001$), and inspection of the upper panel in Figure 3 reveals that this was due mainly to the size of the difference

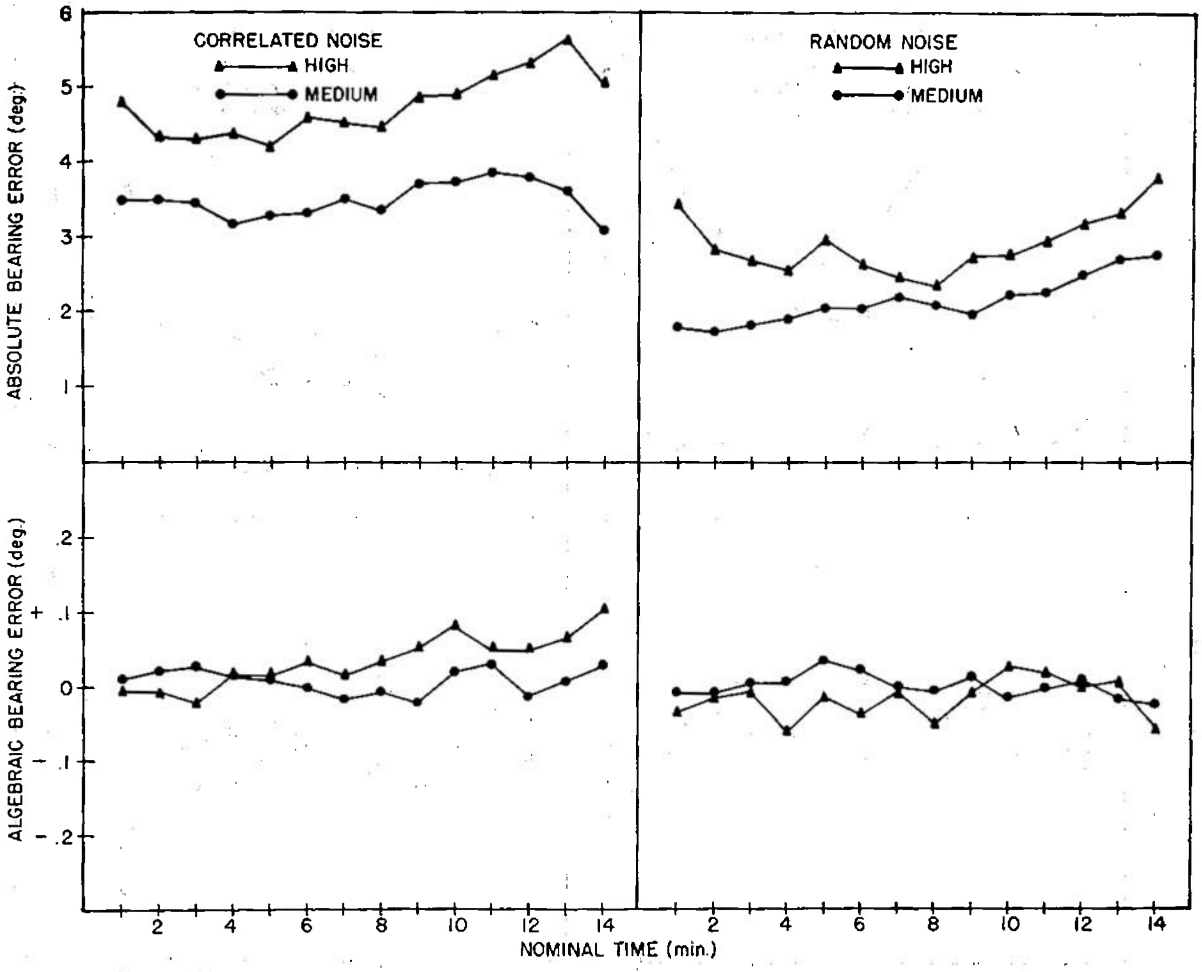

Fig. 4. Error in bearing estimates for Experiment II at two levels of noise.

between the medium and high levels of noise in the two conditions, random and correlated. No such interaction existed in Experiment II. In both experiments performance was reliably different as a function of the hypothetical time in the problem (Exp. I: $F_{13,\ 130} = 2.11$, $p < .05$; Exp. II: $F_{13,\ 104} = 3.42$, $p < .001$). In Experiment I, but not in Experiment II, the interaction between time and level of noise was also significant ($F_{26,\ 260} = 2.56$, $p < .01$). Again, inspection of Figure 3 suggests this was due largely to the contrast of the highly bowed curve for the high-correlated condition with the curve for the high random condition. No other interactions in either experiment were statistically reliable.

Although the algebraic errors indicated that there was no tendency for the subjects' average bearing estimates to deviate from the true bearings, analysis of the absolute errors showed that there were systematic effects on the quality of the bearing estimates as a function of both the level and the kind

of noise. This was true for subjects who saw only one kind of noise, as in Experiment I, or those who saw both kinds, as in Experiment II. There was no systematic tendency for level and kind to interact in any way suggestive of a psychologically meaningful process. The one departure from additivity shown in Figure 3 which contributed to the significant interaction of level and kind with absolute errors in Experiment I has no plausible explanation.

In addition, performance was consistently and reliably different as a function of where the subject was in the problem. Trend analysis confirmed that the effect of time on absolute errors was largely due to inferior performance at the ends of the plots. (All trend analyses in this report are based on the methods discussed in Winer, 1971[1], sec. 7.6, using orthogonal polynomials to partition the main effects into unique trend components.) In Experiment I there were both significant linear and quadratic components (linear: $\underline{F}_{1,130} = 4.02$, $\underline{p} < .05$; quadratic: $\underline{F}_{1,130} = 19.85$, $\underline{p} < .001$). These two components accounted for 87% of the variance due to time. In Experiment II there were significant linear and quadratic trends (linear: $\underline{F}_{1,104} = 27.75$, $\underline{p} < .001$; quadratic: $\underline{F}_{1,104} = 9.64$, $\underline{p} < .01$) and these accounted for 84% of the variance due to time. Inspection of the data in Figures 3 and 4 suggests that the most parsimonious way of describing these trends is to say that performance was most affected at the beginnings and ends of problems. This would be expected since subjects had fewer data points on which to base their estimates of initial and final bearings than in the middle of a plot. This difference between the ends and the middle is highly significant in operational contexts. When the plots are done in real time, most estimation takes place near the end of a continuously growing curve.

The estimation of bearings consists of a combination of tasks beginning with the initial plotting of the raw bearings and ending with the recording of the faired bearings (in these experiments) on a data sheet. Consistent with earlier findings, inspection of the raw data in these experiments indicates that the effects produced by the independent variables were not due to systematic errors in either plotting raw bearings or in reading of faired bearings and entering them on the data sheets. Thus, the locus of the effects is in the actual fairing or subjective curve-fitting that the subject engages in. A full discussion of the human subject as a subjective curve-fitter must await the analysis of the capabilities of objective, mathematical curve-fitting with the present problems. This discussion will be presented later in this report.

Estimation of Bearing Rates

The performance of subjects at estimating bearing rates in these experiments are shown in Figures 5 and 6. Although algebraic errors, shown in the center panel of each figure, lie near zero, there were some tendencies in the data which were confirmed as reliable by statistical analysis. In both experiments the level of the noise was a significant factor (Exp. I: $\underline{F}_{2,20} = 5.02$, $\underline{p} < .05$; Exp. II: $\underline{F}_{1,8} = 9.28$, $\underline{p} < .05$). However, inspection of the center panels in both figures reveals

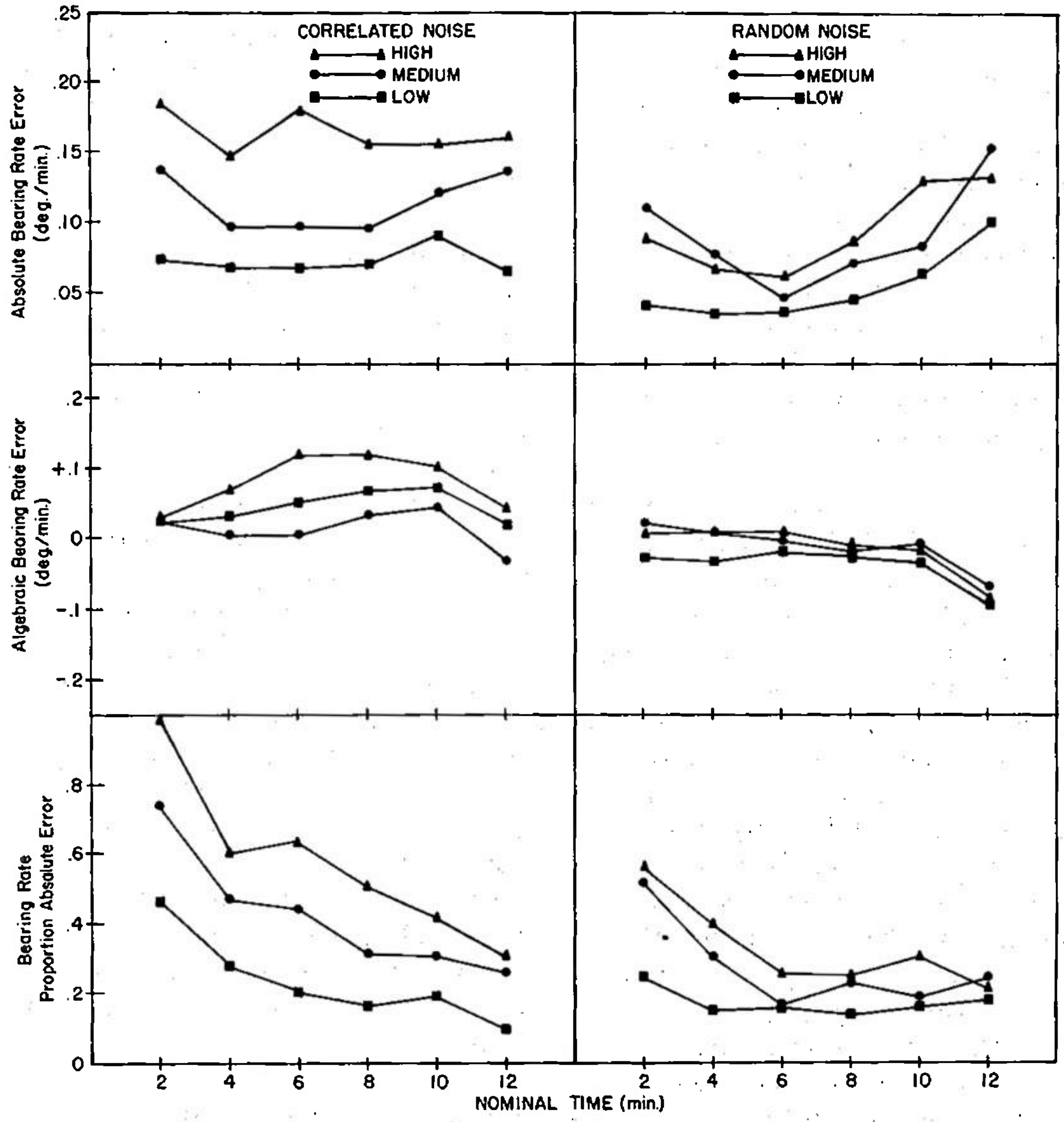

Fig. 5. Error in bearing rate estimates for Experiment I at three levels of noise.

that there is no consistent ordering of the curves as a function of level of noise, making this effect difficult to interpret. In sum, there was at least no straightforward bias in the direction of the average errors. In Experiment I, kind of noise also produced a significant effect on the algebraic errors ($\underline{F}_{1,10} = 3.40$, $\underline{p} < .05$) and kind interacted with level ($\underline{F}_{2,20} = 5.07$, $\underline{p} < .05$). In general, correlated bearings led to estimates of bearing rates which tended to slightly exceed the actual bearing rates, while random bearings produced estimates slightly less than the actual bearing rates. There was also a significant effect in Experiment I due to the time at which the bearing rate was estimated ($\underline{F}_{5,50} = 2.53$, $\underline{p} < .05$), and trend analysis confirmed that this

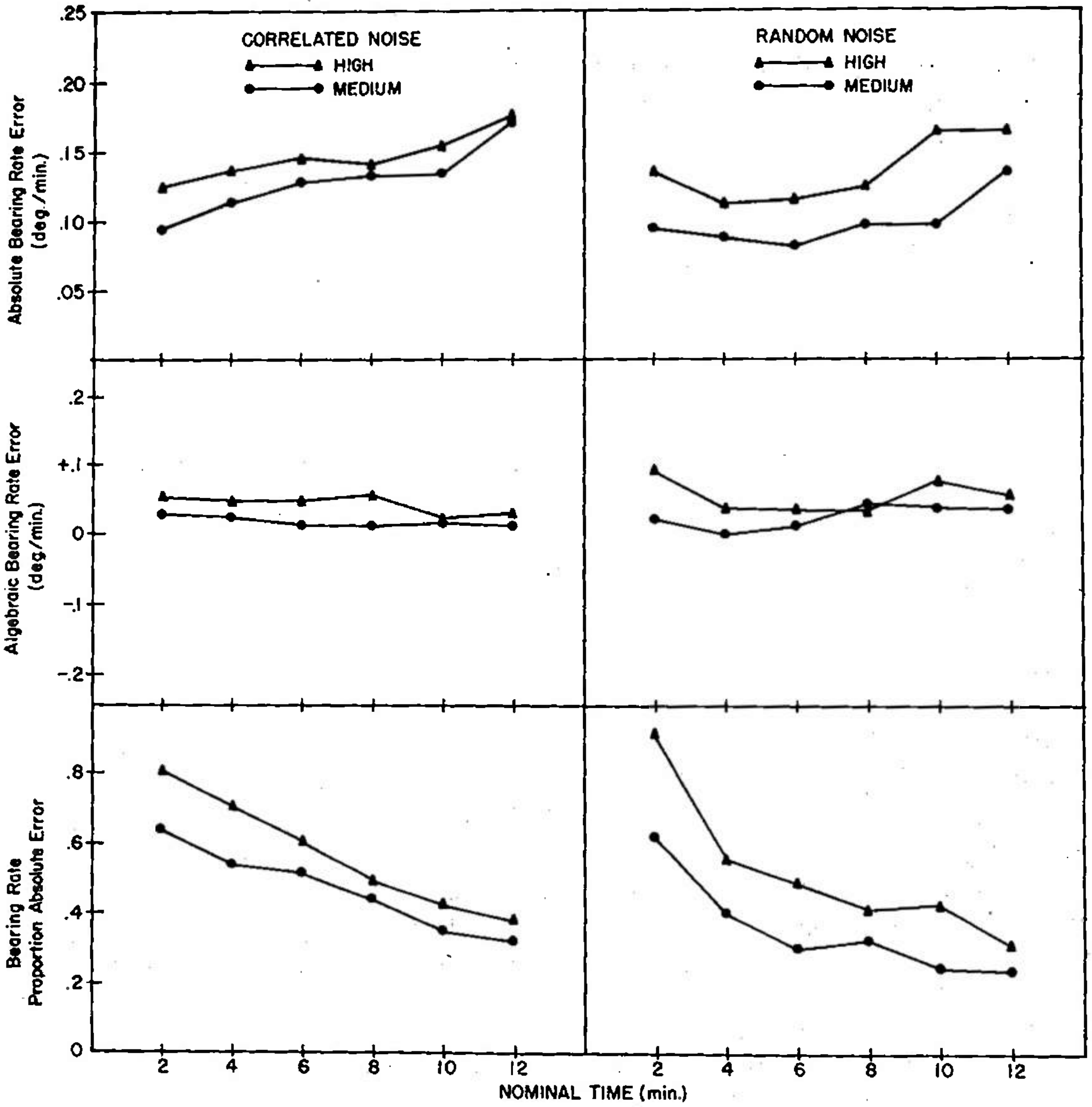

Fig. 6. Error in bearing rate estimates for Experiment II at two levels of noise.

was largely due to the tendency for the estimated bearing rate to drop below the actual rate at the ends of problems. The test for quadratic trend was significant, $\underline{F}_{1,50} = 7.27$, $\underline{p} < .01$. No other trend components reached significance.

The effects due to level and kind of noise on algebraic errors were small, somewhat marginal, and not very systematic, and as a result are exceedingly difficult to interpret. The falling off in the curves in Figure 5 at the last time segment is the most significant effect in terms of the actual plotting task, since it indicates a tendency to underestimate the bearing rates as these rates increase (as they tended to do at the end of most problems).

Further, it indicates some difficulty in estimating the bearing rate at the end of the faired curve, and this is the most realistic condition in terms of real-time performance of this task at sea. However, again this effect was relatively small in magnitude and marginal, and did not show up in Experiment II, so should be viewed with caution.

The upper panels of Figures 5 and 6 show performance as measured by the absolute difference between the actual bearing rate and the estimated bearing rate. As with the bearings, these curves indicate that both the kind and the level of noise affected the quality of the estimates. Larger average deviations were associated with correlated noise as opposed to random, and higher as opposed to lower levels. However, the effect of kind of noise reached statistical significance only in Experiment II (Exp. I: $\underline{F}_{1,10} = 1.21$, n.s.; Exp. II: $\underline{F}_{1,8} = 6.57$, $\underline{p} < .05$) while the effect of level of noise reached significance only in Experiment I. (Exp. I: $\underline{F}_{2,20} = 13.81$, $\underline{p} < .001$; Exp. II: $\underline{F}_{1,8} = 4.97$, $.05 < \underline{p} < .10$). The only other reliable effect in either experiment was the time of estimate, and this was significant in both experiments (Exp. I: $\underline{F}_{5,50} = 3.56$, $\underline{p} < .01$; Exp. II: $\underline{F}_{5,40} = 6.57$, $\underline{p} < .001$). No other effects or interactions reached acceptable levels of statistical reliability.

Unlike the data shown in the center panels for algebraic errors, the data for absolute errors revealed that each of the three independent variables had systematic, orderly effects on the average quality of the estimated bearing rates. The effects for kind and level of noise were not as consistently reliable as with the bearing data considered earlier, but appear nontheless to be important effects. The effect of time of estimate was reliable in both experiments. Trend analysis indicated that the effect of time of estimate could largely be attributed to the relatively more severely degraded performance at the beginning and end of problems (Experiment I: linear component, $\underline{F}_{1,50} = 5.50$, $\underline{p} < .05$, quadratic component, $\underline{F}_{1,50} = 11.60$, $\underline{p} < .01$, the linear and quadratic components accounting for 96% of the variance due to time; Experiment II: linear component, $\underline{F}_{1,40} = 27.89$, $\underline{p} < .001$, quadratic component, $\underline{F}_{1,40} = 4.99$, $\underline{p} < .05$, the linear and quadratic components accounting for 99% of the variance due to time). The ends of these curves represent those portions of the problems where the bearing rate was, in general, either lowest or highest. However, the effect of magnitude of the bearing rate on the accuracy of bearing rate estimation must be examined explicitly before this can be discussed in detail. The ends also represent, of course, those portions of the time bearing curve where the contextual information is minimal.

The lower panels in Figures 5 and 6 present the data in a third format. Here the absolute difference between the actual and estimated bearing rates have been divided by the actual bearing rate, translating an absolute measure into a relative one. The ordinates on these panels represent the amount of error in estimations of the bearing rates as a proportion of the actual underlying bearing rate.

Level of noise systematically ordered these curves, the curve for a higher

level above that of a lower level. This was reliable in both experiments (Experiment I: $F_{2,20} = 24.95$, $p<.001$; Experiment II: $F_{1,8} = 7.28$, $p<.05$). Kind of noise reliably affected this measure in Experiment II ($F_{1,8} = 6.42$, $p<.05$). In Experiment I, kind of noise did not significantly affect proportion absolute error ($F_{1,10} = 2.11$, n.s.) although the direction of the overall means was the same as in Experiment II. Nothing very reliable emerged from this set of measures with respect to the distinction between correlated and random noise, but level of noise consistently affected the percentage measure in both experiments.

This measure also differed reliably as a function of time of estimate (Experiment I: $F_{5,50} = 34.52$, $p<.001$; Experiment II: $F_{5,40} = 19.21$, $p<.001$). Inspection of the lower panels in Figures 5 and 6 shows that this was due to the systematic decrease in proportion absolute error with time in the problem. Trend analysis revealed that both the linear and quadratic trends were significant in both experiments (Experiment I: linear component, $F_{1,50} = 136.07$, $p<.001$, quadratic component, $F_{1,50} = 26.17$, $p<.001$, the linear and quadratic components accounting for 94% of the variance due to time; Experiment II: linear component, $F_{1,40} = 88.24$, $p<.001$, quadratic component, $F_{1,40} = 5.64$, $p<.05$, the linear and quadratic components accounting for 97% of the variance due to time). One major reason for this effect, of course, is that average actual bearing rates were quite small early in the problems, so that small absolute deviations would make for large proportional deviations at these times. Average bearing rates for the six times at which estimates were obtained were: .25, .30, .34, .40, .48, .68. The interaction of time with kind attained significance in both Experiment I ($F_{5,50} = 4.90$, $p<.01$) and in Experiment II ($F_{5,40} = 2.60$, $p<.05$).

Inspection of the data in all panels of Figures 5 and 6 reveals the following picture of human performance at estimating bearing rates in simple time bearing plots. Subjects appear to have no overall bias to their estimates. That is, the expected value of their distributions of estimates appears to be the actual bearing rate, and this expected value is not altered by variations of the independent variables in these experiments. The only exception to this generalization was found in Experiment I, where correlated noise led to significantly larger estimated bearing rates than did random noise. The quality of the estimates as measured by the absolute errors was significantly and systematically affected by each of the independent variables. Increasing the amount of noise or changing from random to correlated noise increased the magnitude of the subjects' absolute errors. Similarly, performance was poorer at the beginnings and ends of problems than in the middle. Subjects apparently used the added contextual information found in the center of the curves to obtain better estimates of the tangent to the curve. In actual operational plots estimates of bearing rate are usually obtained for the last four to six points plotted, and the present data reveal that this is where subjects have the greatest difficulty.

Comparison of Human Performance to that of Rational Curve-Fitting Techniques

In order to evaluate the performance of human subjects on these tasks it is necessary to have some rational measure of the "best" possible performance. The convention adopted here is to contrast the performance of an orthogonal polynomial curve-fitting routine with that of the human subjects in both experiments. The best fitting curves provide a rational optimum for faired bearings, and the derivatives with respect to time (dB/dT) are an optimum estimate of the bearing rate.

Experiments I and II were replicated exactly with a computerized orthogonal polynomial curve-fitting routine substituted for the human subjects. Solutions were restricted to third degree polynomials. Analyses of variance identical to those reported in the last sections were performed, using both the human and computer data, with a between-subjects factor for the source of the data (computer or human) added to the design. The results of these expanded analyses were examined for significant main effects due to the source of the estimates and interactions between source and other independent variables. Consistent trends among the interactions might reveal that there were systematic differences in the character of the solutions provided by each of the two data-generating sources.

Bearing data. Figures 7 and 8 present the data for the computer estimates of bearings, and should be contrasted with the data of Figures 3 and 4. Analysis of variance confirmed that there was a significant main effect of source on absolute errors in both experiments (Experiment I: $F_{1,20} = 8.78$, $p<.01$; Experiment II: $F_{1,16} = 24.73$, $p<.001$), with the performance of human subjects inferior to that of the computer curve-fitting routine. No similar differences were found for algebraic errors. Thus, although there was no difference in the average signed error in the faired bearings of the two data sources, the average quality of the estimates provided by the curve fitting routine was consistently superior to that of the subjects. In other words, subjects' performance did not measure up to the best possible performance.

There were no consistent trends in the interactions of source with the other independent variables, although there were several significant F-ratios. In Experiment I the interaction of level and source for absolute errors was reliable ($F_{2,40} = 4.13$, $p<.05$), while for Experiment II the interaction of kind and source was significant for absolute error ($F_{1,16} = 16.91$, $p<.001$) and the interactions of kind, level, and source ($F_{1,16} = 4.97$, $p<.05$) and time and source ($F_{13,208} = 5.05$, $p<.001$) were significant for algebraic errors. Since no pattern emerged from these, little that is meaningful can be said about the reasons for these interactions. Figure 9 shows in summary form the comparison of the two data sources averaged over level and kind of noise. The significant interaction of time and source for algebraic errors in Experiment II can be clearly seen in this figure. Why the curve for the computer generated data should differ so markedly from that for the human data is unclear. The cause may lie in a

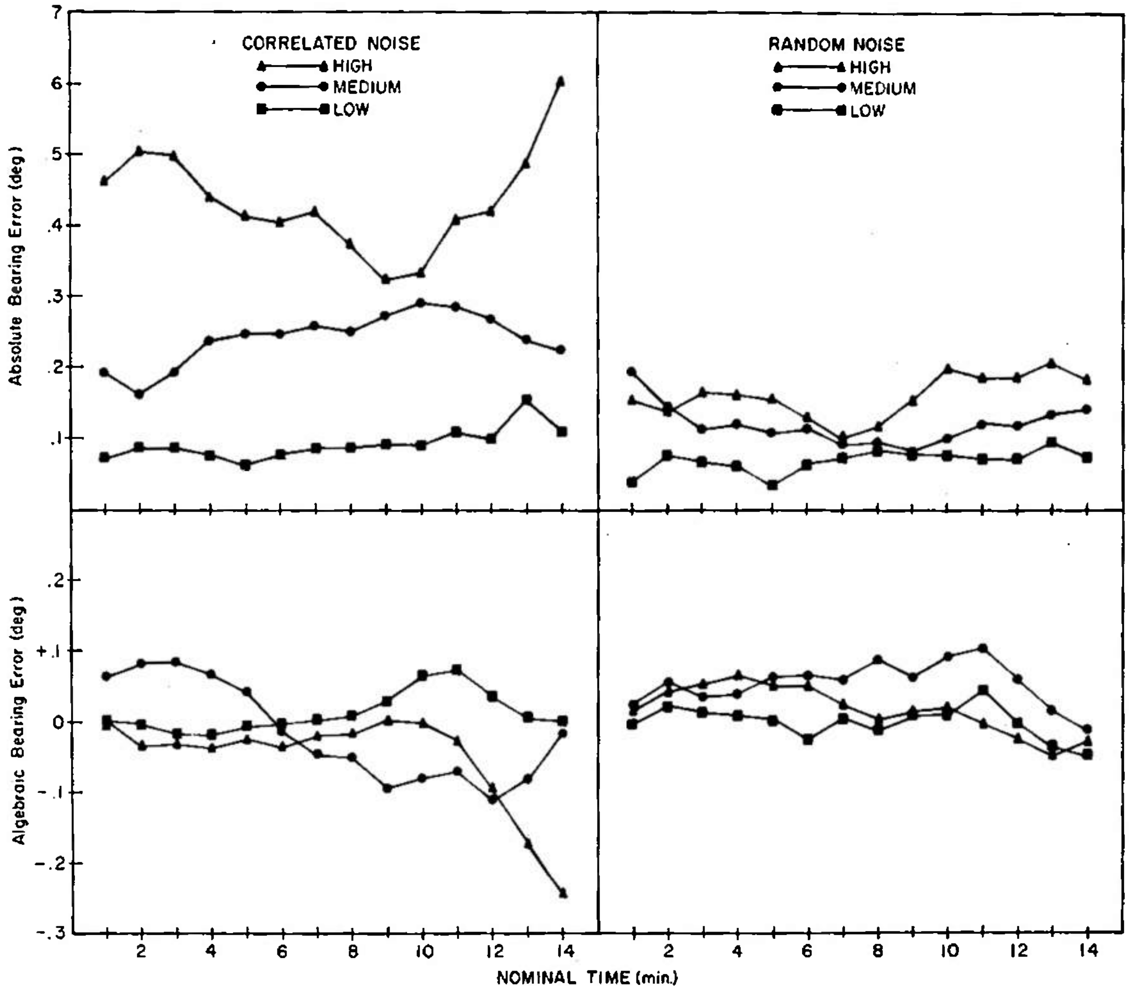

Fig. 7. Error in computer bearing estimates for Experiment I at three levels of noise.

peculiar relationship of the weaknesses of this particular curve-fitting routine with the subset of problems in Experiment II. Since no similar relationship emerged in Experiment I, it cannot be attributed to any general characteristics of the curve-fitting routine.

Bearing rate data. Figures 10 and 11 contain the data relevant to the performance of the curve-fitting routine, and Figures 5 and 6 have the comparable data for human subjects. The curve-fitting routine was a better estimator of bearing rates than were the human subjects on two measures: absolute error (Experiment I: $F_{1,20} = 4.36$, $p < .05$; Experiment II: $F_{1,16} = 13.68$, $p < .01$) and proportion absolute error (Experiment I: $F_{1,20} = 4.83$, $p < .05$; Experiment II: $F_{1,16} = 4.74$, $p < .05$). Algebraic errors were not reliably different for the two data sources.

The only systematic pattern of interactions to emerge were those involving source and time. Figure 12

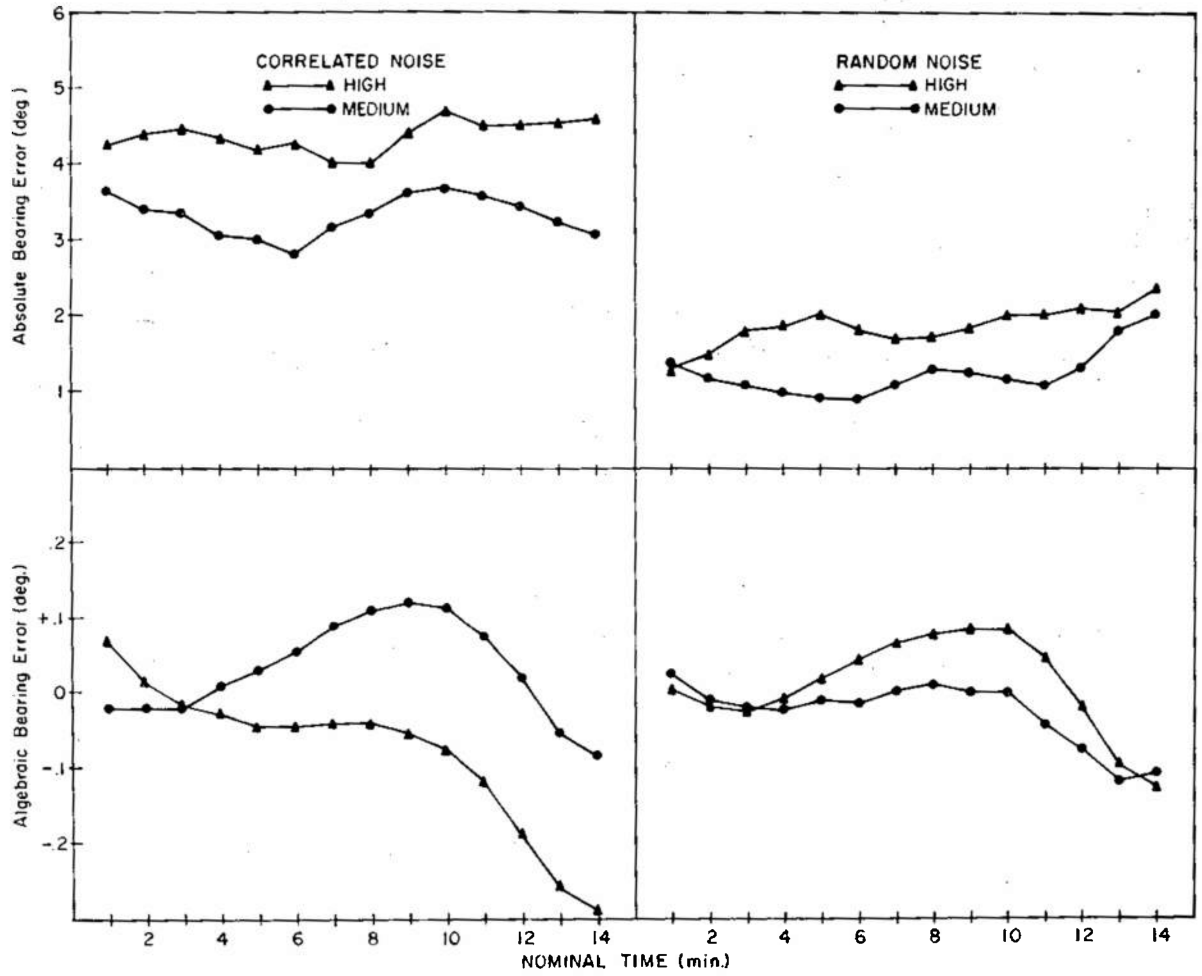

Fig. 8. Error in computer bearing estimates for Experiment II at two levels of noise.

summarizes the bearing rate data and shows the contrast of computer-generated data and human data averaged over kind and level of noise. The interactions reached significance on the following measures: Experiment I, algebraic error ($\underline{F}_{5,100} = 2.84$, $\underline{p} < .05$) and proportion absolute error ($\underline{F}_{5,100} = 4.97$, $\underline{p} < .001$); Experiment II, absolute error ($\underline{F}_{5,80} = 4.07$, $\underline{p} < .01$). These interactions indicate that nominal time affected the performance of the two data sources differently. Where these interactions attained significance it appears from Figure 12 that this was due to the computer-generated data yielding a more severely bowed function than the human data. Table 2 summarizes the tests on the differences in trends that contribute to the source by time interactions, and shows that the curves differed in either their linear or quadratic trends. Thus, although the overall performance of the computer routine at estimating bearing rates was superior to that of humans,

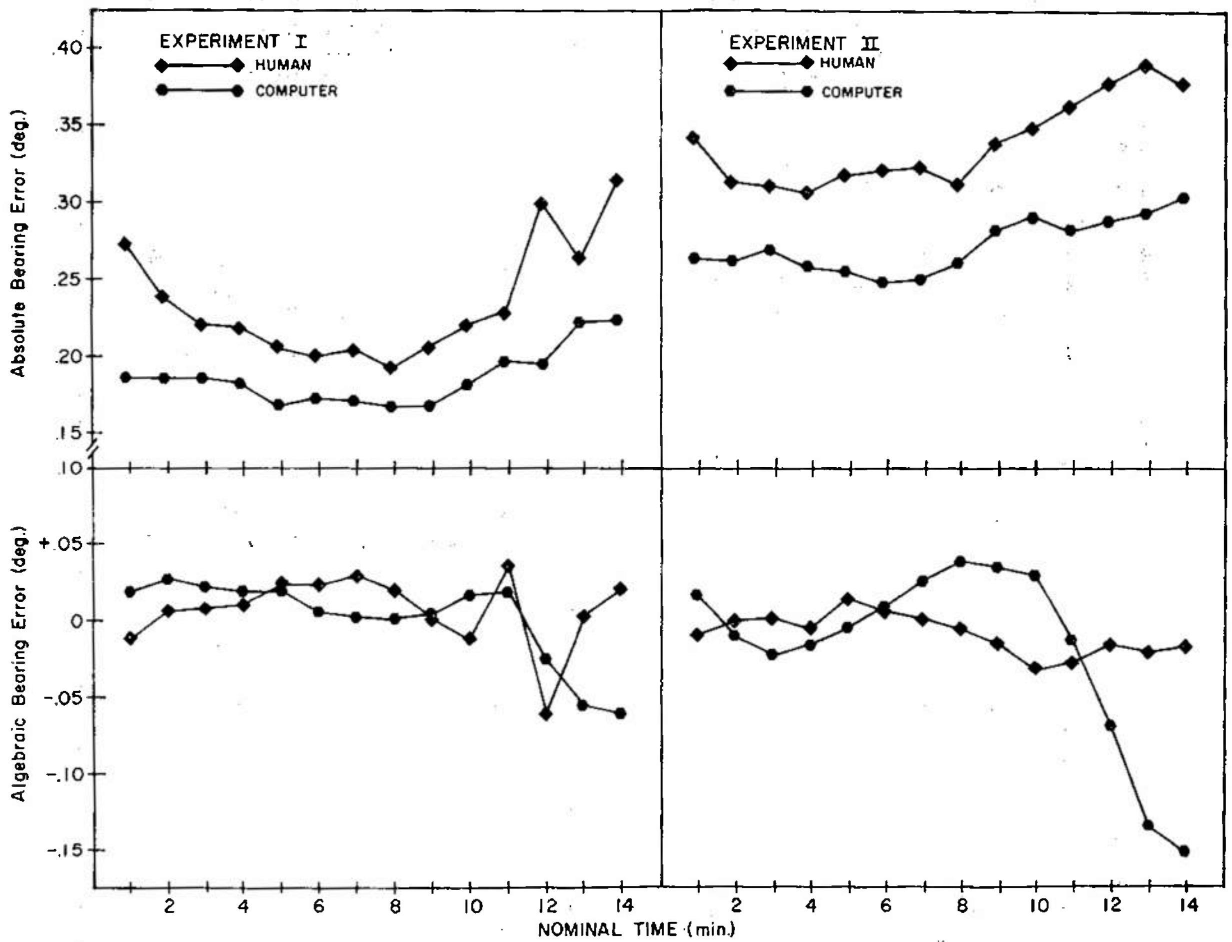

Fig. 9. Summary comparisons of human and computer error in bearing estimates.

the former had relatively more difficulty capturing the bearing rates of the end points of the time bearing curves.

Summary of comparison. The time bearing performance of human subjects was consistently inferior to that of a mathematical estimator. This was true for both the fairing of bearings and the estimating of bearing rates. Figures 9 and 12 make clear, however, that although many of the differences were reliable, they were relatively small in size, especially for the estimation of bearings. With bearing rates, the errors of subjects were large relative to the computer estimates but were still small in absolute magnitude.

All of the independent variables except time had the same kinds of effects on the performance of both data sources. Thus, the fact that correlated noise yields poorer estimates of bearings and bearing rates is due to the characteristics of the data, not those of the

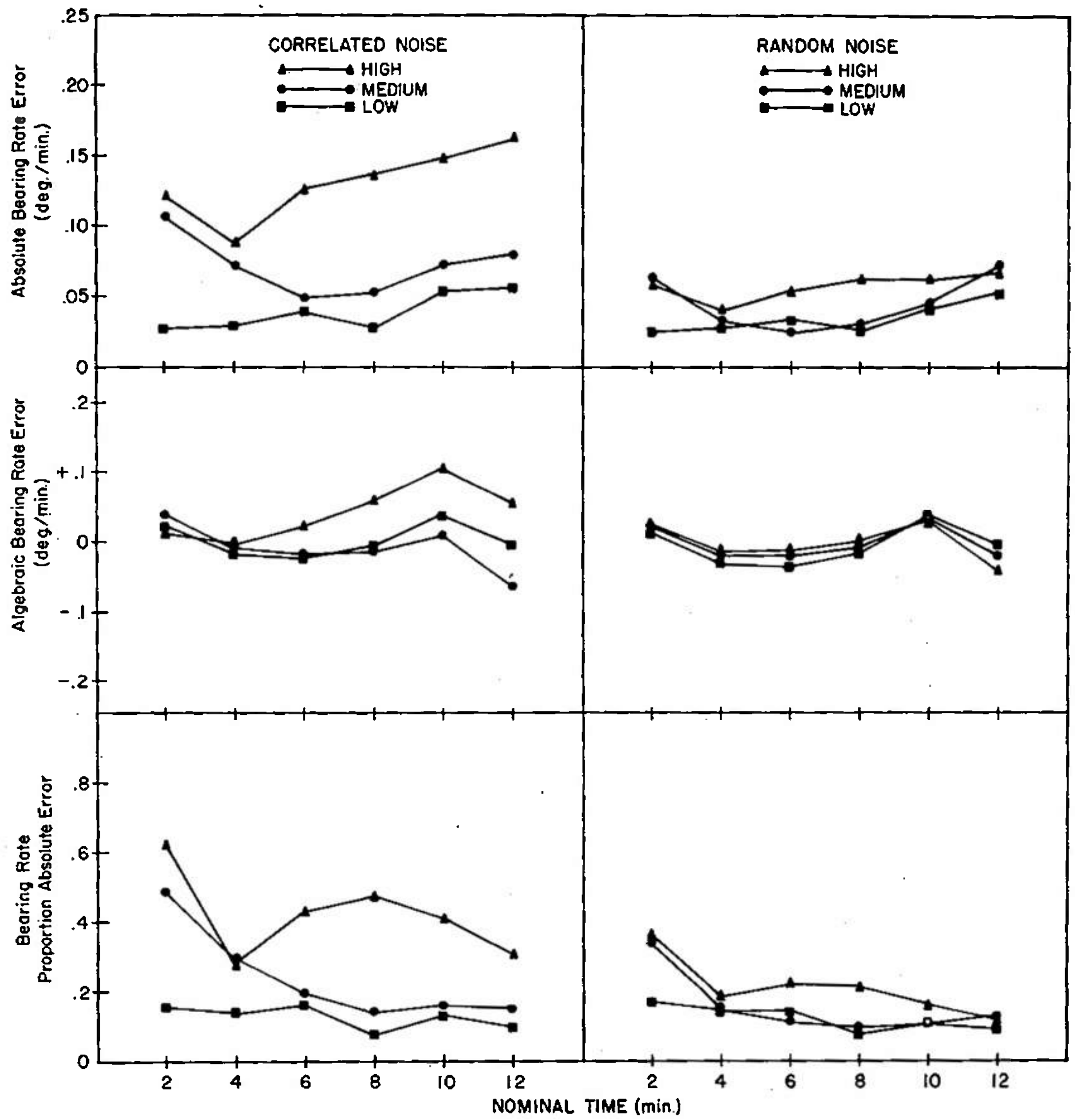

Fig. 10. Error in computer bearing rate estimates for Experiment I at three levels of noise.

human estimator. The lack of consistent interactions between either kind or level of noise and source of data indicates that these properties of the raw bearings have effects on the estimates that cannot be eliminated by altering the way in which the human does his task. Although the analyses reported earlier in this report indicated that subjects had difficulties at the ends of time bearing curves in estimating both bearings and bearing rates, the analyses just discussed indicate that, if anything, the curve-fitting routine has even more trouble. Thus, the human operator appears to be an especially effective integrator of information at the ends of curves like these and

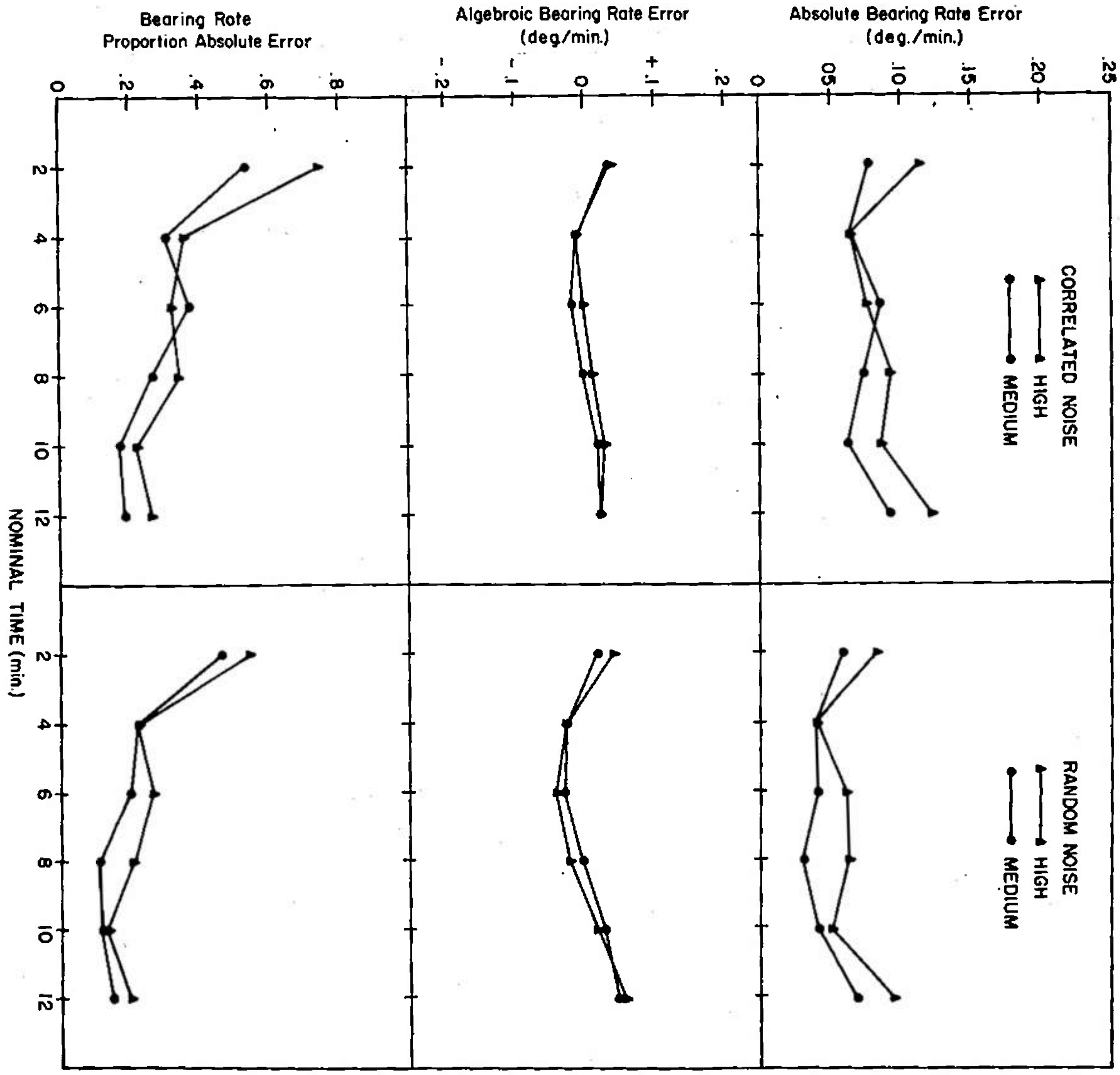

Fig. 11. Error in computer bearing rate estimates for Experiment II at two levels of noise.

produces estimates of bearings and bearing rates that are less distorted by the lack of a context than does the mathematical estimator. This is encouraging since in operational plots the estimation usually takes place near the end of a growing curve in real time.

Performance on Problem 4.

As noted earlier, Problem 4 had to be excluded from the main analysis because it yielded performance radically different from the other eleven problems. Figure 1 shows why. Problem 4, one

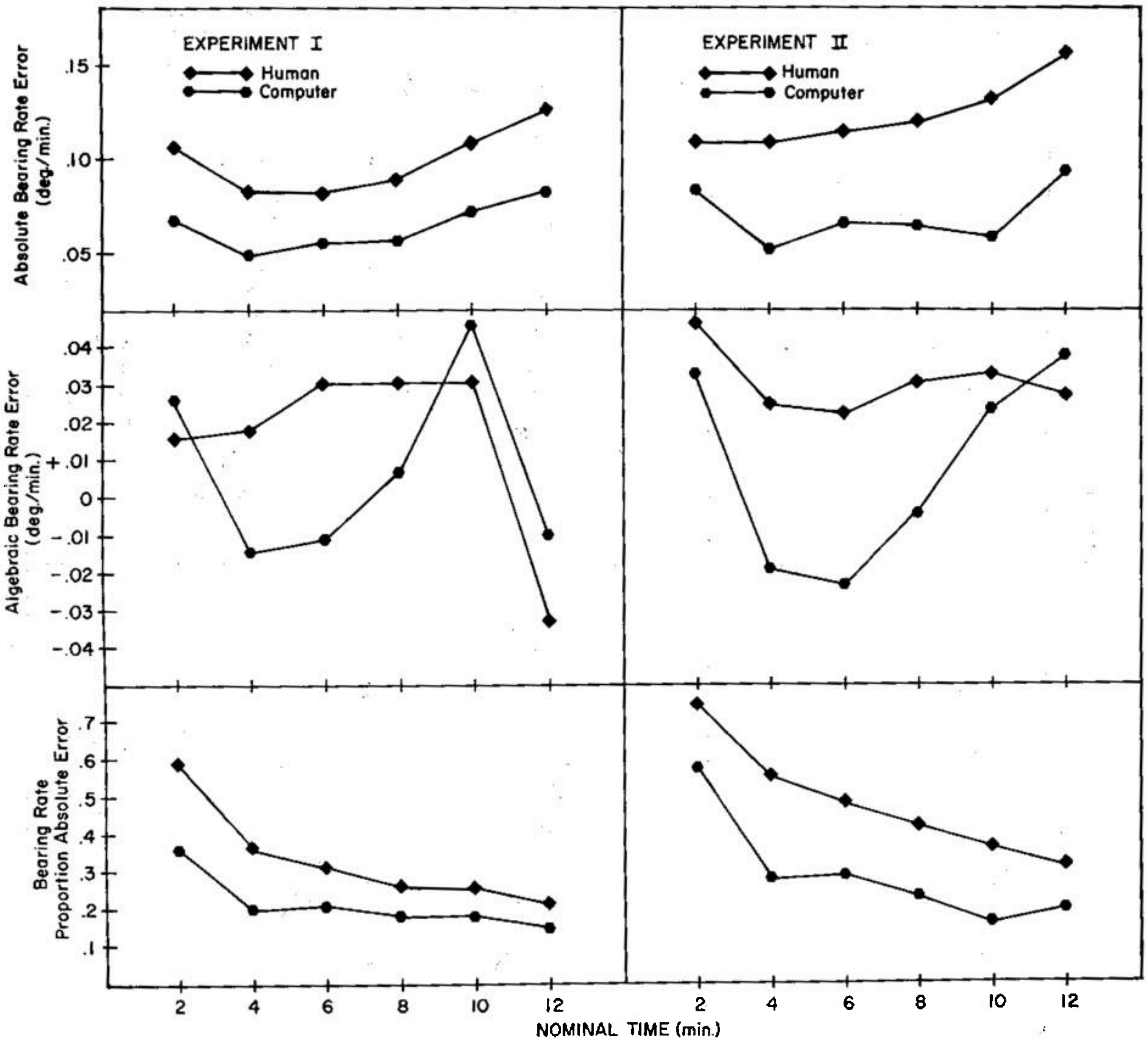

Fig. 12. Summary comparisons of human and computer error in bearing rate estimates.

of those composed of line segments, passed through an inflection point. That is, there was a region of very high bearing rate preceded and followed by regions of low bearing rate. This would be typical of a time bearing curve produced by a target or own ship maneuver, or by passage through the point of minimum range (CPA) for fixed target and own ship courses. Thus, it represents a situation which frequently arises in operational situations.

The character of Problem 4 caused great difficulty for both the human performers (recall that Problem 4 was used only in Experiment I) and for the analytic curve-fitting routine, apparently because the noise obscured the large change in bearing rate.

Table 2. Summary of Trend Components For Source X Time Interactions in Bearing Rate Estimates

	F Linear	F Quadratic	Proportion of Variance Accounted for by Linear + Quadratic
Exp. I: Algebraic Error $df = 1, 100$	3.08	8.58*	83%
Exp. I: Proportion Absolute Error $df = 1, 100$	20.52**	3.71	98%
Exp. II: Absolute Error $df = 1, 80$	11.08*	1.75	66%

* $p < .01$
** $p < .001$

Both subjects and the computer routine seriously underestimated the bearing rate at nominal time six minutes, that time when the bearing rate was very high (it was 2.00, contrasting with .40 at time 4 and .20 at time 8). Figures 13 through 17 summarize the performance of the two data sources at the various time intervals for bearings and bearing rates. These data have been collapsed over kind and level of noise, since insufficient data existed (each subject saw Problem 4 only once, at one level and one kind of noise) for a complete analysis. Only time in the problem and the interaction of time with the source of the data (human or machine estimations) yielded consistently significant results and these are the data of greatest interest to us here.

Figures 13 and 14 show that the analytic curve fitting routine had considerably greater difficulty estimating bearings for Problem 4 than did human estimators. Since the routine was set to find the best third degree solution using orthogonal polynomials, in principle it should have been able to recover a function like Problem 4. However, the routine seriously overestimated the bearings at time 4, 5,

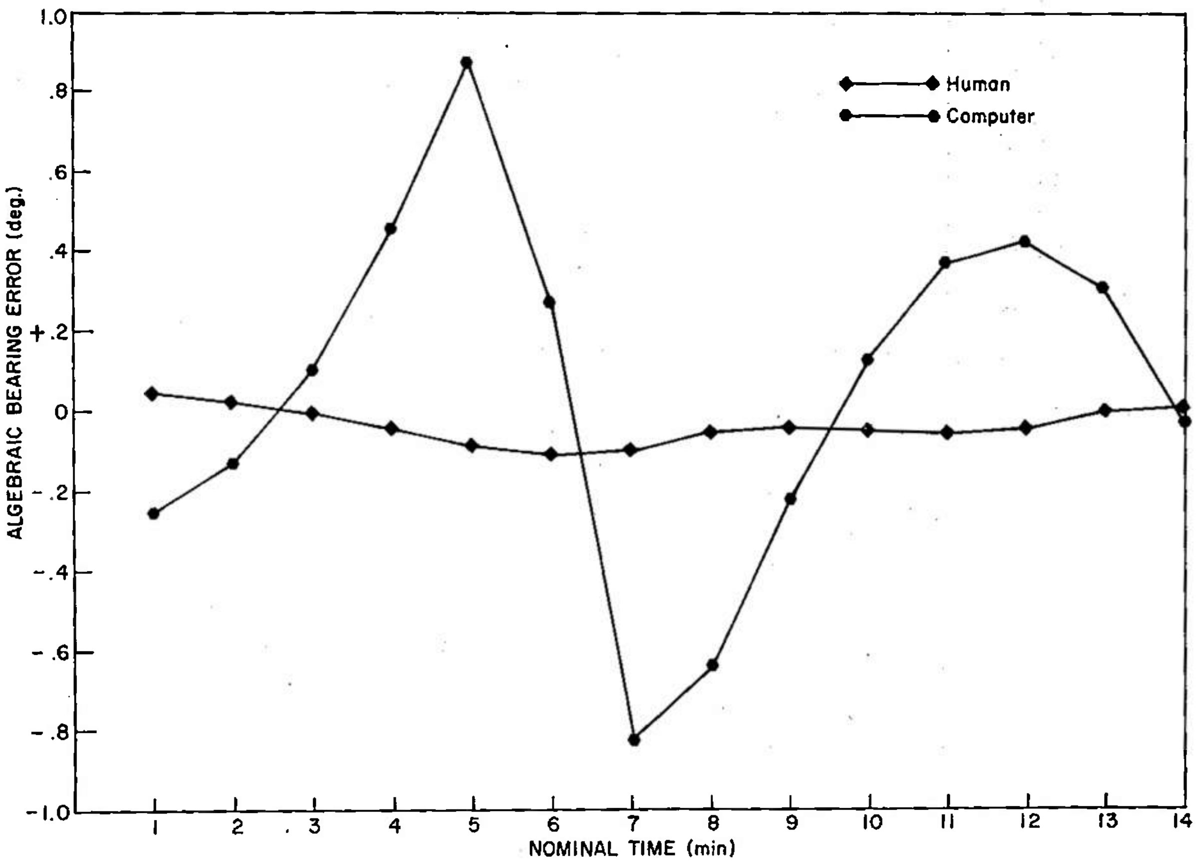

Fig. 13. Human and computer algebraic error in bearing estimates for Problem 4.

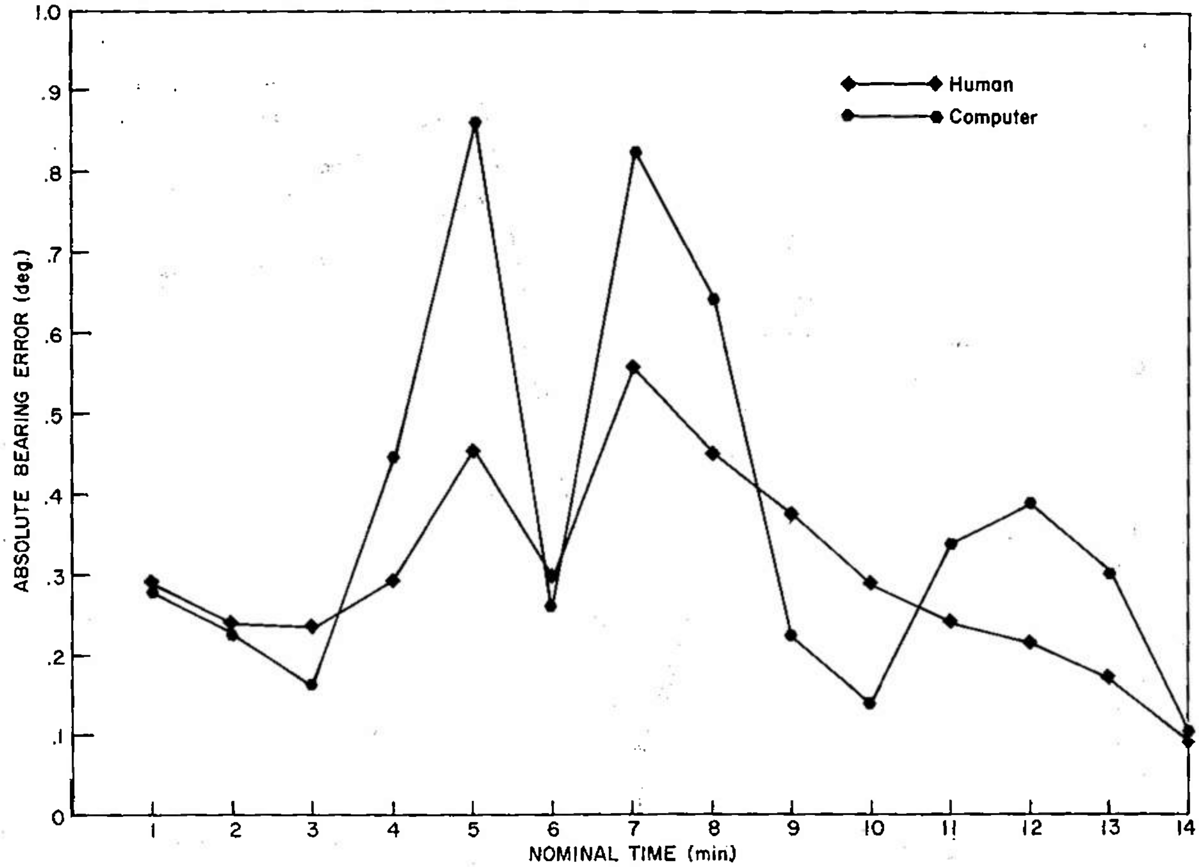

Fig. 14. Human and computer absolute error in bearing estimates for Problem 4.

and 6, and seriously underestimated those at times 7, 8, and 9. Bearings were again overestimated at times 10 through 13. Human performers' average algebraic errors were consistently very close to zero, although Figure 12 shows that they too had difficulty at the times listed above.

The data of Figures 15-17 show that although the forms of the various functions are fairly similar for human and mathematical estimators, the mathematical estimator is consistently inferior. Statistical analysis confirmed that this was significant for each of the measures of bearing rate (algebraic error: $\underline{F}_{1,12} = 7.30$, $\underline{p} < .05$; absolute error: $\underline{F}_{1,12} = 11.51$, $\underline{p} < .01$; proportion absolute error: $\underline{F}_{1,12} = 20.34$, $\underline{p} < .001$). The main effect for time was of course significant for all measures (algebraic error: $\underline{F}_{5,60} = 223.46$, $\underline{p} < .001$; absolute error: $\underline{F}_{5,60} = 81.15$, $\underline{p} < .001$; proportion absolute error: $\underline{F}_{5,60} = 144.92$, $\underline{p} < .001$), but so was the interaction of time with human vs. analytic estimators (algebraic error: $\underline{F}_{5,60} =$

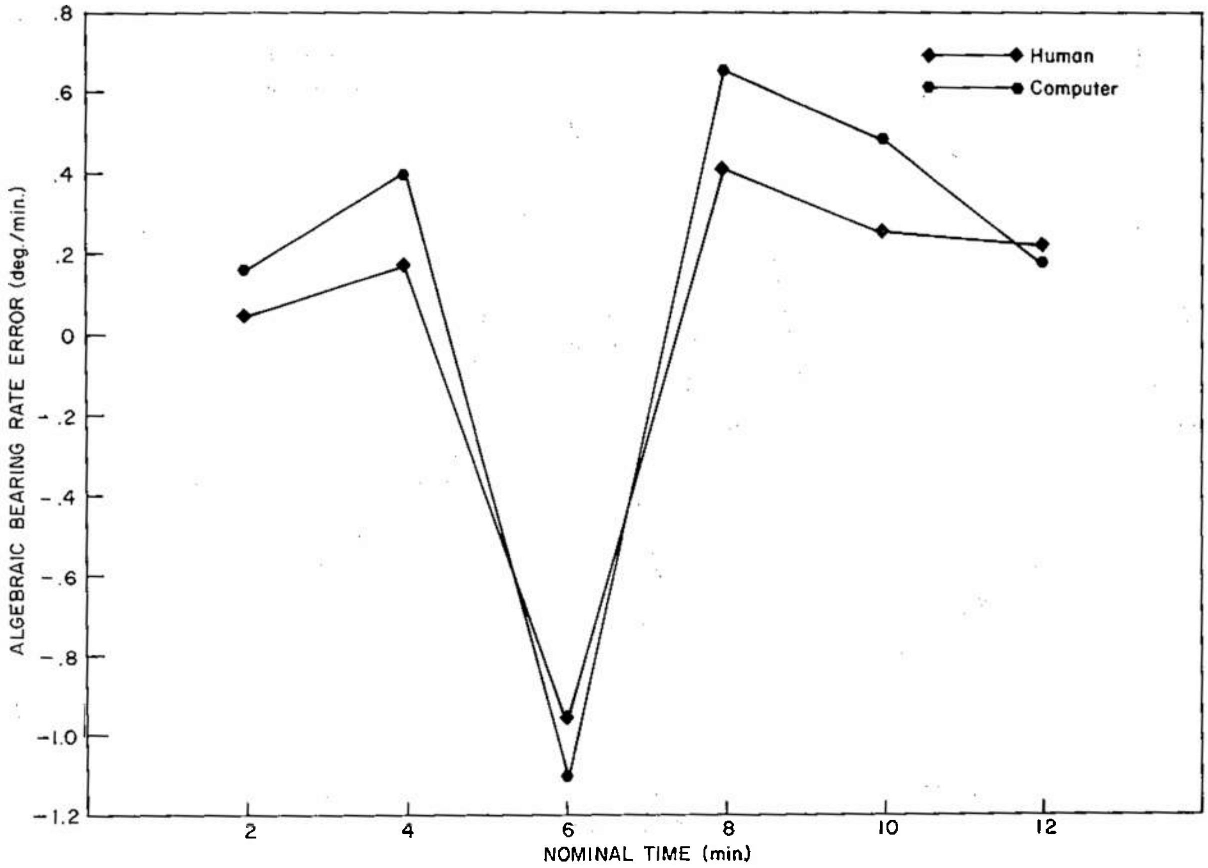

Fig. 15. Human and computer algebraic error in bearing rate estimates for Problem 4.

4.67, $\underline{p}<.01$; absolute error: $\underline{F}_{5,60}$ = 2.48, $\underline{p}<.05$; proportion absolute error: $\underline{F}_{5,60}$ = 15.88, $\underline{p}<.001$).

The trend of these data for both bearings and bearing rates is just the opposite that of the data for the other eleven problems. In the analysis reported in the previous section, the performance of the mathematical curve fitting routines was consistently superior to that of the human estimators. On Problem 4 the human subjects were consistently superior. With a sample of only one problem of this kind we are hesitant to make confident generalizations, but since the effects were so statistically reliable it leads us to believe the difference is real. Garnatz and Hunt (1971)[2] have reported differences in the relative ease with which human estimators and mathematical estimators can find solutions in tasks somewhat different than this, but their results along with those reported here suggest there are general classes of estimation problems for which the human's perceptual and cognitive

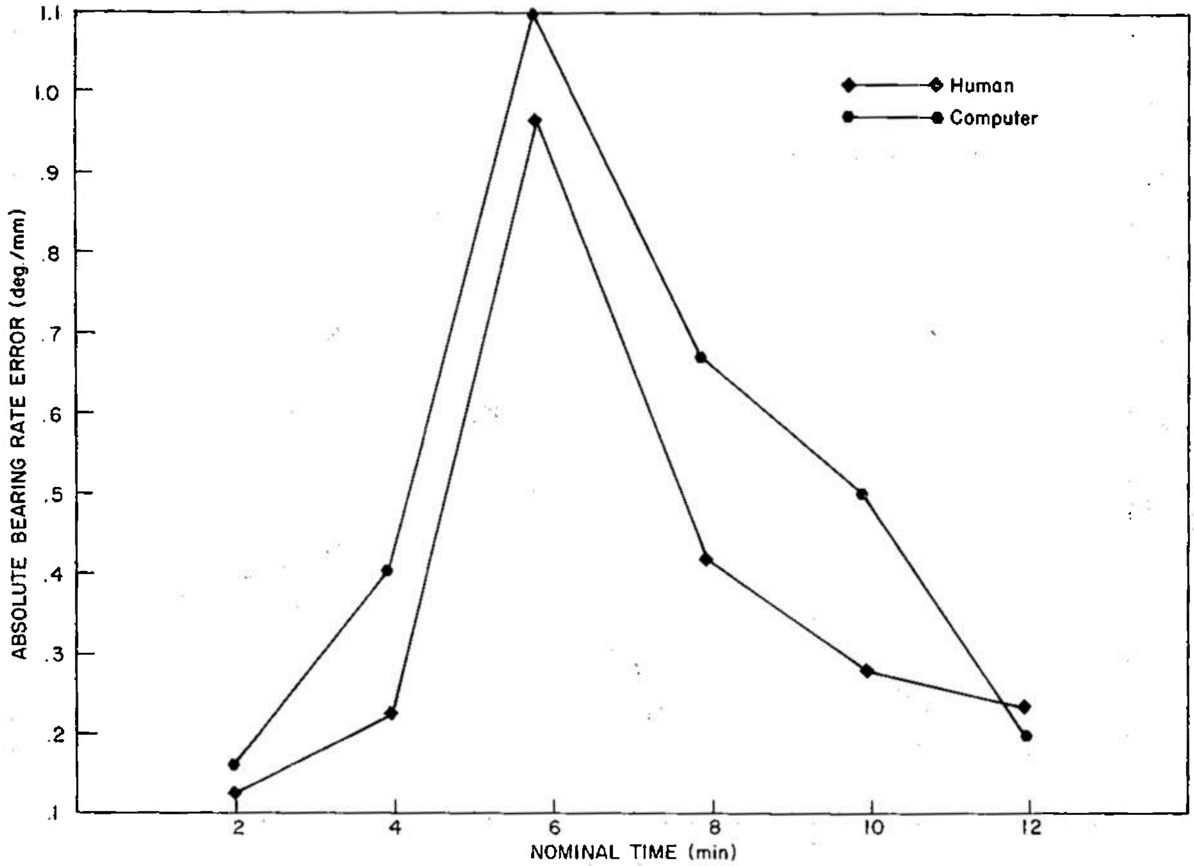

Fig. 16. Human and computer absolute error in bearing rate estimates for Problem 4.

abilities give him a distinct advantage over simple mathematical estimators.

Comparison of Experienced vs. Inexperienced Subjects in Experiment II.

Two quite different populations of subjects were represented in Experiment II. On the one hand there were four civilians recruited from a local program for the disadvantaged, and on the other there were five Naval enlisted men, all with some pre-experimental exposure to the time bearing plot and to fire control operations. An auxiliary analysis was carried out to see if this difference in background resulted in performance differences. One of the five Navy men was excluded from the analysis since he differed in age and experience from the other four (he was an E-7, while the other four were E-4's).

The mean performance curves indicated that the civilian subjects were consistently inferior to the military subjects. Since analysis of

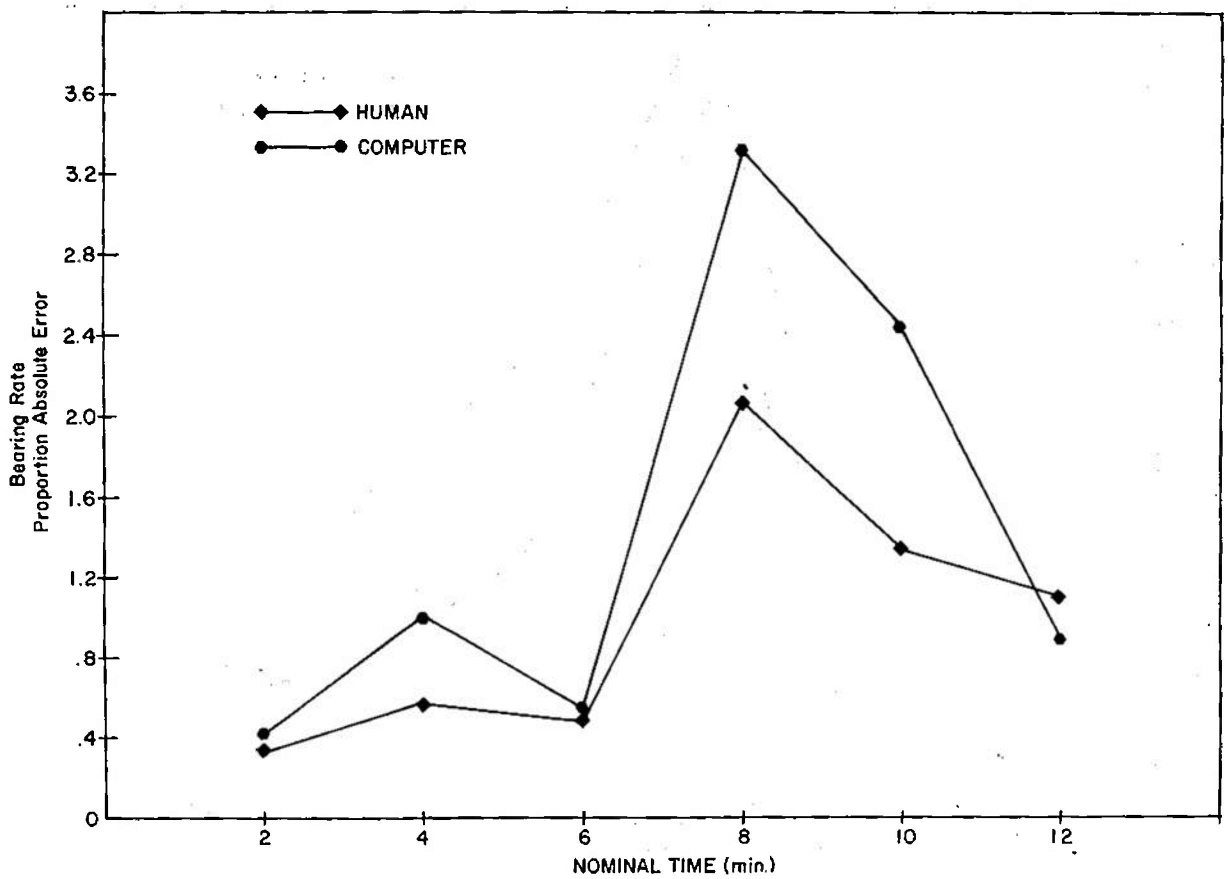

Fig. 17. Human and computer proportion of absolute error in bearing rate estimates for Problem 4.

variance on the data for the eight subjects would have been based on only four observations per cell, a non-parametric test was performed comparing means from civilian and military subjects for each nominal problem time within each noise condition. Tables 3 and 4 show these means and the significance levels obtained using the Sign Test (Siegel, 1956[3], pp. 68-75) for bearing estimates and bearing rate estimates, respectively.

In both cases, comparing mean algebraic errors, only the high level correlated noise showed a significant difference. In the cases of mean absolute errors, however, all differences were significant except the high level correlated noise condition. For bearing rate estimates, all conditions of proportion absolute error showed significant differences. Apparently the military subjects were able to make use of their previous experience in performance on this task.

Table 3. Comparison of Bearing Estimates From Civilian and Military Subjects in Experiment II and Sign Test Significance Levels

	Mean Bearing Estimate Error		Level of Significance
	Civilian	Military	
Algebraic Error			
Correlated Noise			
High	.097	-.062	$\underline{p}$ = .046
Medium	-.006	.017	n.s.
Random Noise			
High	-.003	-.002	n.s.
Medium	.031	-.030	n.s.
Absolute Error			
Correlated Noise			
High	.481	.474	n.s.
Medium	.396	.321	$\underline{p}$ = .006
Random Noise			
High	.360	.234	p < .001
Medium	.245	.197	$\underline{p}$ = .001

Table 4. Comparison of Bearing Rate Estimates From Civilian and Military Subjects in Experiment II and Sign Test Significance Levels

Mean Bearing Rate Estimate Error			Level of Significance
	Civilian	Military	
Algebraic Error			
Correlated Noise			
High	.080	.016	p = .016
Medium	.023	.012	n.s.
Random Noise			
High	.060	.045	n.s.
Medium	.025	.021	n.s.
Absolute Error			
Correlated Noise			
High	.169	.133	n.s.
Medium	.143	.116	$\underline{p}$ = .016
Random Noise			
High	.167	.107	p = .016
Medium	.118	.074	$\underline{p}$ = .016
Proportion Absolute Error			
Correlated Noise			
High	.739	.447	p = .016
Medium	.549	.382	$\underline{p}$ = .016
Random Noise			
High	.718	.359	p = .016
Medium	.466	.248	$\underline{p}$ = .016

SUMMARY

From the variety of results obtained in these experiments we can draw the following conclusions about the questions investigated. The kind and level of noise in the raw bearings have a systematic and consistent effect on the quality of performance in the time bearing plot, although they do not differentially affect the average algebraic direction of the errors in the subject's estimates. That is, there is no tendency for subjects to either consistently overestimate or underestimate time bearing outputs as a function of the properties of the noise. Furthermore, since an analytic curve-fitting technique was affected in the same way by the properties of the noise, there is no way to eliminate these effects short of eliminating the noise itself from the data. The direction of the bearing rate had no effect at all on the performance of subjects estimating bearings or bearing rates.

An interesting pattern emerged from the comparisons of the human estimator with the mathematical one. Overall human performance was consistently inferior to that of the curve-fitting routine, but the magnitude of the differences was not large. The curve-fitting routine had relatively more trouble with the ends of the time bearing curves, and on the one problem that contained a more complicated curve shape, it was inferior overall to the human estimator. Thus, although the human cannot do as well as the curve-fitting routine when the shapes of the curves are simple, when the estimation must be done with relatively little context or with more complicated curves, the human estimator produces more stable results than the curve-fitting routine. The human's ability to apply cognitive constraints to his task and to exploit perceptual abilities unavailable to the mathematical device is apparently of great benefit.

A subsidiary analysis indicated that experience with time bearing problems yielded better performance, although the gains were small. However, in Experiment II the change in performance over a 30-day confinement was negligible. The longer-term effect was probably due to a greater understanding of the role of the time bearing task in fire control problems.

REFERENCES

1. Winer, B.J. Statistical principles in experimental design. (2nd ed.) New York: McGraw-Hill, 1971.

2. Garnatz, D., & Hunt, E. Eyeball parameter estimation with a computer. Technical Report No. 71-11-08, Computer Science Group, University of Washington, November 1971.

3. Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

4. Feller, W. An introduction to probability theory and its applications. Volume 1. (3rd ed.) New York: Wiley, 1968.

5. Green, B.F., Jr. Digital computers in research. New York: McGraw-Hill, 1963.

Generation of the Statistical Noise

The central limit theorem of probability theory states that the sum of n independent random variables with a common distribution approaches a normal distribution as n tends to infinity (see, e.g., Feller, 1968)[4]. It can be shown that n = 12 is a satisfactory approximation for many applications, and that if the n random variables are real numbers on the interval (0, 1), with n = 12 the distribution of the sums will have unit variance (Green, 1963[5], pp. 170-171). This principle was used to write a computer routine to generate the statistical noise for the raw bearings. To create the random noise, 12 numbers were generated using a subroutine whose output was a pseudo-random number from a uniform distribution on the interval (0, 1). These were summed, and the constant six was subtracted from the sum, yielding a pseudo-random number from a normal distribution with zero mean and unit variance. The result of 300 samples using the routine for random noise is shown in Figure A-1. Correlated noise was generated in essentially the same way, except that each new sum consisted of 11 of the pseudo-random uniform numbers from the previous sum plus one new one. (The random noise, of course, used twelve new numbers for each new sum.) The

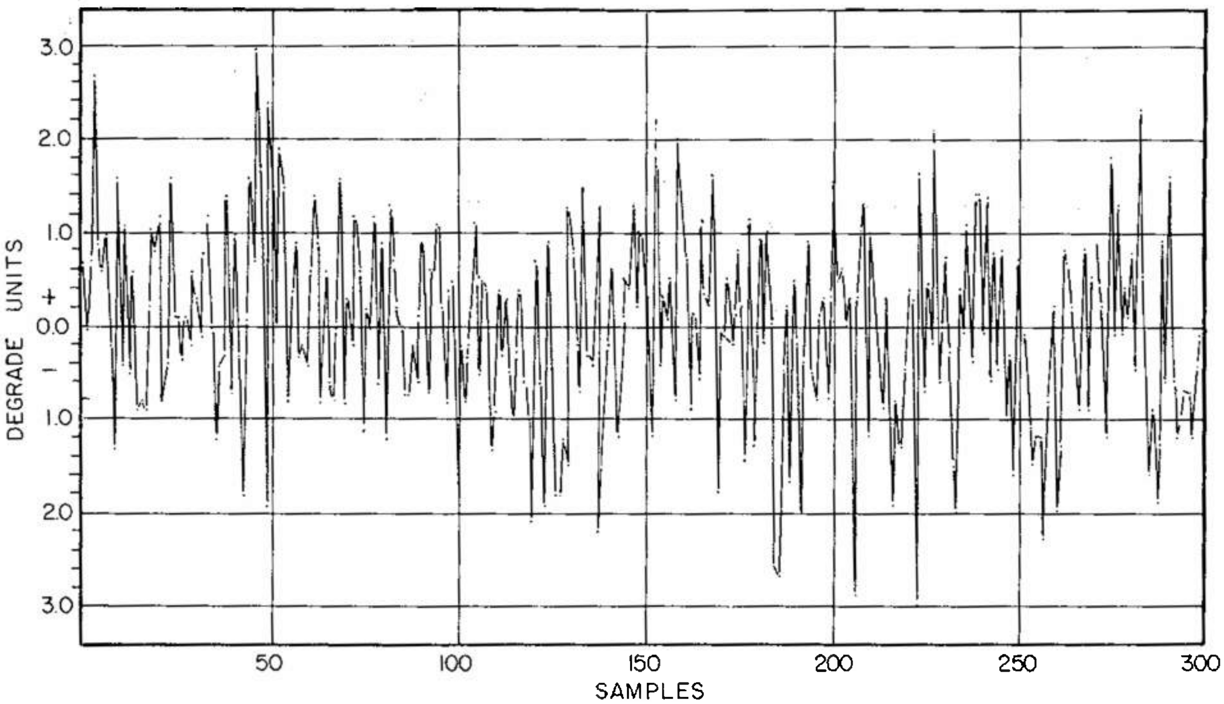

Fig. A-1. Three hundred successive samplings of random noise.

output of this process was also adjusted to have a zero mean, and examples of 300 successive samplings are shown in Figure A-2. Level of noise was controlled by multiplying the adjusted sum by a scale factor calculated in accordance with standard sonar equations to simulate conditions that might be found in operations at sea. In sum, the raw bearing was computed according to the following equation:

raw bearing = actual bearing + (adjusted sum) X (scale factor)

The adjusted sum is, of course, distributed around zero.

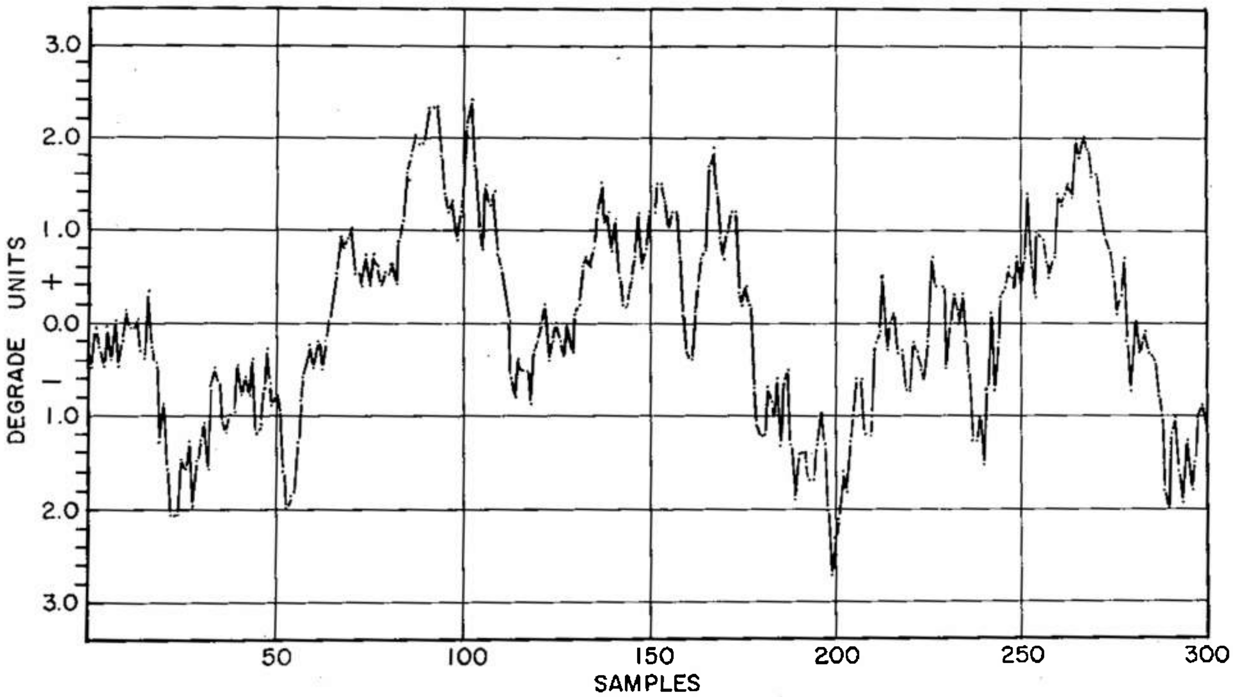

Fig. A-2. Three hundred successive samplings of correlated noise.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Naval Submarine Medical Research Laboratory Naval Submarine Medical Center	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

PERFORMANCE ON THE EXPANDED TIME BEARING PLOT AS A FUNCTION OF BEARING ACCURACY

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Interim report

5. AUTHOR(S) (First name, middle initial, last name)

GARY M. OLSON
KEVIN LAXAR

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
26 June 1972	30 and 1 Appendix (2 pp)	11

8a. CONTRACT OR GRANT NO. b. PROJECT NO. MF51.524.004.2002D c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) NSMRL Report Number 716 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Submarine Medical Center Box 600 Naval Submarine Base Groton, Conn. 06340

13. ABSTRACT

Two experiments analyzed the effects of statistical noise in raw sonar bearings on performance in a laboratory version of the expanded time bearing plot. Accuracy of faired bearings and bearing rate estimates were taken as the measures of performance. Greater amounts of noise led to poorer performance, but these decrements were smaller when the noise was random than when it was correlated. Human performance was contrasted with that of an orthogonal polynomial curve fitting routine designed to do the same task. The mathematical routine was affected by the noise in the same way as humans were. However, on simple plots the mathematical routine provided superior solutions while on curves of more complex shapes or at the ends of curves humans were superior. Thus, in certain situations the human's perceptual and cognitive abilities gave him a distinct advantage over the mathematical routine.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Time Bearing Plot Bearing Degradation Human Information Processing Analytic Curve-fitting Faired Bearings Bearing Rate Estimation Fire Control						